

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א



$$\{\sqrt{x}\}^2$$



תוכן העניינים

1. סטטיסטיקה תיאורית- הצגה של נתונים 1
2. סטטיסטיקה תיאורית- סכימה 9
3. סטטיסטיקה תיאורית- מדדי מרכז (ללא ספר) 9
4. סטטיסטיקה תיאורית- מדדי פיזור (ללא ספר) 13
5. סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני 17
6. סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן 19
7. סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה 21
8. סטטיסטיקה תיאורית-מקדם ההשתנות 23
9. סטטיסטיקה תיאורית שאלות אמריקאיות 30
10. מקדם המתאם (מדד קשר) הלינארי ומובהקותו (ללא ספר)
11. מדד הקשר של ספירמן (ללא ספר)
12. התפלגויות רציפות מיוחדות - התפלגות נורמלית (ללא ספר) 41
13. הסקה סטטיסטית - הקדמה 44
14. התפלגות הדגימה ומשפט הגבול המרכזי 50
15. מושגי יסוד באמידה 55
16. מבוא לבדיקת השערות על פרמטרים 61
17. רווח סמך לתוחלת (ממוצע) 69
18. בדיקת השערות על תוחלת (ממוצע) 69

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 1 - סטטיסטיקה תיאורית- הצגה של נתונים

תוכן העניינים

1. כללי 1

סטטיסטיקה תיאורית – הצגה של נתונים:

רקע:

דרכים להצגת נתונים שנאספו:

רשימה של תצפיות:

התצפית היא הערך שנצפה עבור ישות מסוימת בקבוצה. רושמים את התצפיות שהתקבלו כרשומה, יעיל שיש מספר מועט של תצפיות. ההצגה הזו רלבנטית לכל סוגי המשתנים. למשל, להלן מספר החדרים בבניין בן 5 דירות: 3, 4, 3, 5, 4.

טבלת שכיחויות בדידה:

שם המשתנה- X	שכיחות – $f(x)$	שכיחות יחסית באחוזים
X_1	f_1	$\frac{f_1}{N} \cdot 100$
X_2	f_2	$\frac{f_2}{N} \cdot 100$
X_3	f_3	$\frac{f_3}{N} \cdot 100$
\vdots	\vdots	\vdots
X_k	f_k	$\frac{f_x}{N} \cdot 100$
סה"כ	$N = \sum_{i=1}^k f_i$	100%

רושמים את התצפיות בטבלה שבה עמודה אחת מבטאת את ערכי המשתנה והשנייה את השכיחות. יעיל עבור משתנה איכותי וכמותי בדיד וכשיש מספר רב של תצפיות. לא יעיל למשתנה כמותי רציף.

דוגמה:

להלן התפלגות הציונים בכיתה מסוימת:

$\frac{f_i}{n}$	F_i	מספר התלמידים – השכיחות f	הציון X
$0.08=2/25$	2	2	5
$0.16=4/25$	6	4	6
$0.32=8/25$	14	8	7
$0.2=5/25$	19	5	8
$0.16=4/25$	23	4	9
$0.08=2/25$	25	2	10

שכיחות מצטברת – צבירה של השכיחויות.

השכיחויות F_i – השכיחות המצטברת נותנת כמה תצפיות קטנות או שוות לערך.

שכיחות יחסית (פרופורציה) – השכיחות מחולקת לכמות התצפיות הכללי:

$$\frac{f_i}{n} - \text{איזה חלק מהתצפיות בקבוצה שוות לערך.}$$

טבלת שכיחויות במחלקות:

משתמשים שהמשתנה כמותי רציף או כאשר יש מספר ערכים רב במשתנה הבדיד וטבלת שכיחויות תהיה ארוכה מידי.

דוגמה:

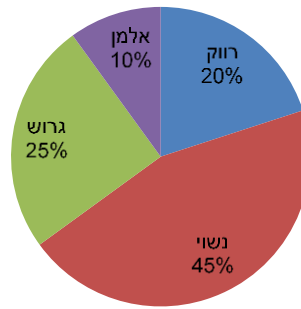
נתנו לקבוצת ילדים לבצע משימה, בדקו את התפלגות זמן הביצוע, בדקות. להלן ההתפלגות שהתקבלה:

מספר הילדים	זמן בדקות
20	0.5-3.5
18	3.5-9.5
14	9.5-19.5
8	19.5-29.5

דיאגרמת עוגה:

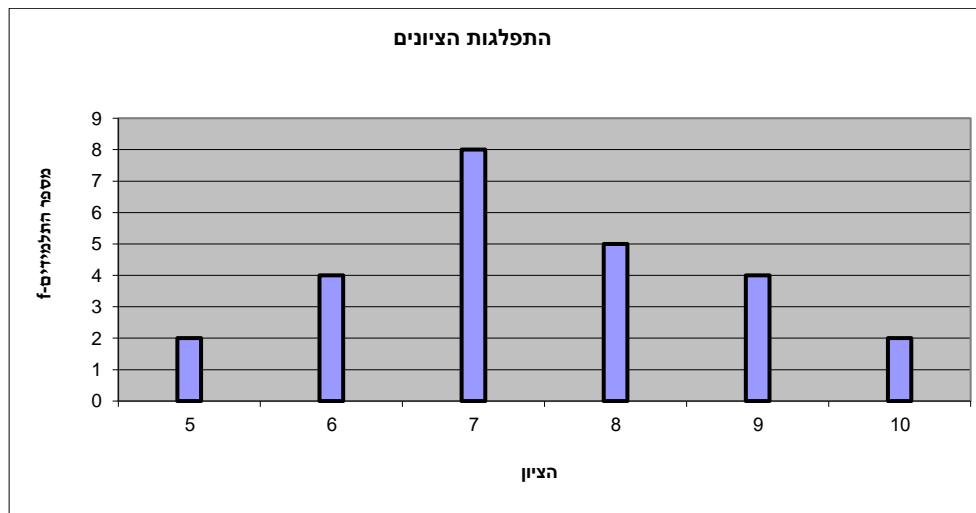
זהו התיאור הגרפי של משתנה איכותי. בדיאגרמת עוגה כל ערך במשתנה מקבל "נתח", שהוא פרופורציונלי לשכיחות היחסית של ערך המשתנה בנתונים.

התפלגות המצב המשפחתי



דיאגרמת מקלות:

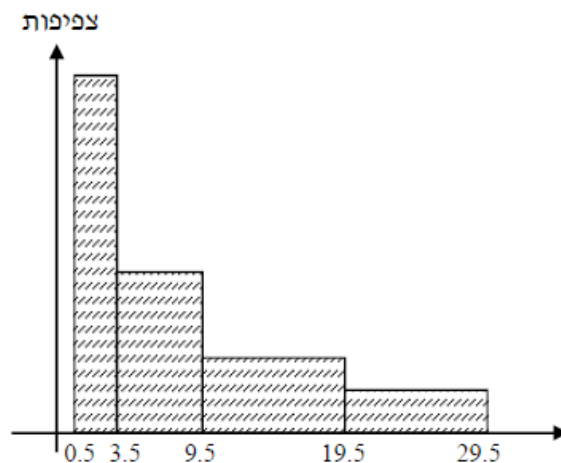
הציר האופקי הוא הציר של המשתנה והציר האנכי של השכיחות, כך שהגובה של המקל מעיד על השכיחות. רלבנטי למשתנה כמותי בדיד. לא נהוג להשתמש בתיאור למשתנה איכותי וכמו כן לא למשתנה כמותי רציף, וכן בסולמות מדידה עבור משתנה מסולם סדר.



היסטוגרמה:

היסטוגרמה היא הדרך הגרפית כדי לתאר טבלת שכיחויות במחלקות, והיא רלוונטית למשתנה כמותי רציף. בהיסטוגרמה הציר האופקי הוא הציר של המשתנה והציר האנכי הוא הציר של הצפיפות. הצפיפות מחושבת בכל מחלקה על ידי חלוקת השכיחות ברוחב של כל המחלקה, והיא נותנת את מספר התצפיות הממוצע בכל מחלקה ליחידה. אם המחלקות הן שוות ברוחב, ניתן לשרטט את ההיסטוגרמה לפי השכיחות ואין צורך בצפיפות.

צפיפות	מצטברת	שכיחות	אמצע	רוחב	X
6.6667	20	20	2	3	0.5 - 3.5
3	38	18	6.5	6	3.5 - 9.5
1.4	52	14	14.5	10	9.5 - 19.5
0.8	60	8	24.5	10	19.5 - 29.5



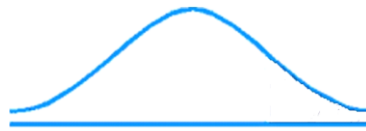
פוליגון – מצולעון:

אם נחבר את אמצע קצה כל מלבן בקווים ישרים. נותן מראה חזותי לצורה של התפלגות המשתנה.

צורות התפלגות נפוצות:

התפלגות סימטרית פעמונית

רוב התצפיות במרכז, וככל שנתרחק מהמרכז יהיו פחות תצפיות באופן סימטרי. לדוגמה, ציוני IQ.

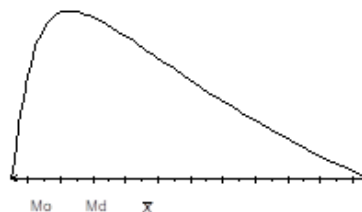


ישנן התפלגויות סימטריות שאינן פעמוניות, כגון:

התפלגות אסימטרית ימנית (חיובית)

רוב התצפיות מקבלות ערכים נמוכים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים גבוהים קיצוניים. לדוגמה, שכר במשק.

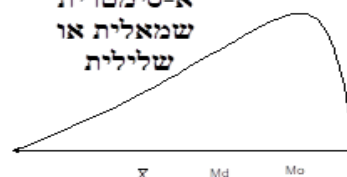
התפלגות א-סימטרית ימנית או חיובית



התפלגות אסימטרית שמאלית (שלילית)

רוב התצפיות מקבלות ערכים גבוהים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים נמוכים קיצוניים. לדוגמה, אורך חיים.

התפלגות א-סימטרית שמאלית או שלילית



שאלות:

- 1) בסקר צפייה בטלוויזיה התקבלו התוצאות הבאות: 25 צפו בערוץ הראשון, 25 צפו בערוץ 10, 75 צפו בערוץ השני, 50 צפו באחד מערוצי הכבלים ו-25 לא צפו בטלוויזיה בזמן הסקר.
- א. רשמו את טבלת השכיחות ואת השכיחות היחסית.
- ב. תארו את הנתונים באופן גרפי.

- 2) להלן נתונים על התפלגות המקצוע המועדף של תלמידי שכבה ו' בבית הספר "מעוף":

המקצוע	מספר התלמידים
מתמטיקה	44
תנ"ך	20
אנגלית	12
היסטוריה	26

- א. מהו המשתנה הנחקר?
- ב. מהי פרופורציית התלמידים שמעדיפים תנ"ך?

- 3) להלן התפלגות ההשכלה במקום עבודה מסוים:

השכלה	מספר העובדים
נמוכה	60
תיכונית	120
אקדמאית	20

- א. מהו המשתנה הנחקר?
מאיזה סולם הוא?
- ב. תארו את הנתונים באופן גרפי.

- 4) להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:
- 6, 5, 8, 7, 6, 8, 9, 6, 7, 6, 7, 8, 5, 8, 4, 6, 10, 9, 8, 6, 7.
- א. מהו המשתנה? האם הוא בדיד או רציף?
- ב. תארו את הרשימה בטבלת שכיחויות.
- ג. הוסיפו שכיחויות יחסיות לטבלה.
- ד. תארו את הנתונים באופן גרפי.

תשובות סופיות:

1) א. להלן טבלה:

ב. עיין גרף מלא בסרטון הוידאו.

%	$\frac{f(x)}{n}$	$f(x)$	x
12.5%	$\frac{25}{200}$	25	ערוץ 1
12.5%	$\frac{25}{200}$	25	ערוץ 10
37.5%	$\frac{75}{200}$	75	ערוץ 2
25%	$\frac{50}{200}$	50	כבלים
12.5%	$\frac{25}{200}$	25	לא צפו
100%	1	200	סה"כ

2) א. מקצוע מועדף.

ב. 19.6%.

3) א. משתנה נחקר: השכלה, סוג: סדר.

ב. עיין גרף מלא בסרטון הוידאו.

ב+ג. להלן טבלה:

%	$\frac{f(x)}{n}$	$f(x)$	x
5%	$\frac{1}{20}$	1	4
10%	$\frac{2}{20}$	2	5
30%	$\frac{6}{20}$	6	6
20%	$\frac{4}{20}$	4	7
20%	$\frac{4}{20}$	4	8
10%	$\frac{2}{20}$	2	9
5%	$\frac{1}{20}$	1	10
100%	20	20	סה"כ

4) א. המשתנה: ציון, משתנה בדיד.
 ד. עיין גרף מלא בסרטון הוידאו.

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 2 - סטטיסטיקה תיאורית- סכימה

תוכן העניינים

1. כללי 9

סטטיסטיקה תיאורית – סכימה:

רקע:

בסטטיסטיקה ישנה צורת רישום מקובלת לסכום של תצפיות: $\sum_{i=1}^n X_i$.

נסביר את צורת הרישום על ידי הדוגמה הבאה:

i	X_i
1	5
2	0
3	1
4	3
5	2

(הסבר מלא מופיע בסרטונים באתר).

שאלות:

- 1) בבניין 5 דירות. לכל דירה רשמו את מספר החדרים שיש בדירה (X), ומספר הנפשות החיות בדירה (Y) חשבו:

Y	X	מספר דירה
1	2	1
1	3	2
2	2	3
3	4	4
2	3	5

א. $\sum_{i=1}^3 X_i$

ב. $\sum_{i=1}^5 Y_i$

ג. $\sum_{i=1}^4 X_i$

ד. $\left(\sum_{i=1}^4 X_i\right)^2$

ה. $\sum X_i$

ו. $\sum X_i Y_i$

ז. $\sum(X_i) \sum(Y_i)$

- (2) נתון לוח ערכי המשתנים X_i ו- Y_i , כאשר: $i = 1, 2, \dots, 6$, ונתונים הקבועים:
 $a = 2$, $b = 5$. חשבו את הנוסחאות הבאות:

i	1	2	3	4	5	6
X_i	3	2	4	-2	1	4
Y_i	2	0	0	1	-5	2

$$\text{א. } \sum_{i=1}^4 y_i$$

$$\text{ב. } \sum_{i=1}^6 a$$

$$\text{ג. } \sum_{i=1}^6 x_i y_i$$

$$\text{ד. } \sum_{i=1}^6 (x_i + y_i)$$

$$\text{ה. } \sum_{i=1}^6 x_i + a$$

- (3) קבעו לכל זהות האם היא נכונה:

$$\text{א. } \sum_{i=1}^n bX_i = b \cdot \sum_{i=1}^n X_i$$

$$\text{ב. } \sum_{i=1}^n a = a \cdot n$$

$$\text{ג. } \left(\sum_{i=1}^n X_i \right)^2 = \sum_{i=1}^n X_i^2$$

תשובות סופיות:

- | | | | |
|---------|--------------|-----------|--------------|
| ד. 121. | ג. 11. | ב. 9. | א. 7. (1 |
| | ז. 126. | ו. 27. | ה. 14. |
| ד. 12. | ג. 7. | ב. 12. | א. 3. (2 |
| | | | ה. 14. |
| | ג. לא נכונה. | ב. נכונה. | א. נכונה. (3 |

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 3 - סטטיסטיקה תיאורית- מדדי מרכז

תוכן העניינים

1. כללי (ללא ספר)

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 4 - סטטיסטיקה תיאורית- מדדי פיזור

תוכן העניינים

1. כללי (ללא ספר)

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 5 - סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני

תוכן העניינים

1. טווח בינרבעוני.....13

סטטיסטיקה תיאורית – מדדי פיזור – טווח בין רבעוני:

רקע:

הטווח הבין-רבעוני (יש הקוראים לו התחום הבין-רבעוני) נותן את הטווח בין הרבעונים בו נמצאים 50% מהתצפיות המרכזיות. הרעיון ליצור מדד פיזורי שלא רגיש לתצפיות החריגות ביותר. כדי לחשב את הטווח הבין-רבעוני יש למצוא את הרבעון התחתון והעליון של התפלגות התצפיות.

רבעון תחתון – ערך שמחלק את ההתפלגות לשניים.
25% מהמקרים נמוכים ממנו או שווים לו ו-75% מהמקרים גבוהים או שווים לו.
סימון: Q_1 .

רבעון עליון – ערך שמחלק את ההתפלגות לשניים.
75% מהמקרים נמוכים ממנו או שווים לו ו-25% מהמקרים גבוהים או שווים לו.
סימון: Q_3 .

הטווח הבין רבעוני הוא הפער בין שני הרבעונים: $IQR = Q_3 - Q_1$.

שלב ב' במציאת טווח בין-רבעוני בטבלת שכיחויות:

שלב א': נמצא את הרבעון התחתון: הוא הערך שהשכיחות היחסית המצטברת באחוזים עברה לראשונה את 25%.

שלב ב': נמצא את הרבעון העליון: הוא הערך שהשכיחות היחסית המצטברת באחוזים עברה לראשונה את 75%.

שלב ג': נמצא את הטווח הבין-רבעוני: נחסר את הרבעונים: $IQR = Q_3 - Q_1$.

דוגמה (פתרון בהקלטה):

בסניף בנק 250 לקוחות. ספרו לכל לקוח את מספר תוכניות החיסכון שלו. מהו הטווח הבין-רבעוני של מספר תוכניות החיסכון בסניף?

שכיחות יחסית מצטברת	שכיחות מצטברת	$f(x)$	# תוכניות החיסכון
		100	0
		75	1
		25	2
		25	3
		25	4

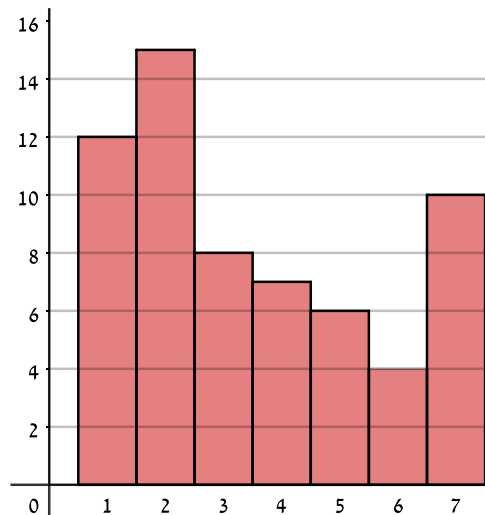
שאלות:

1) להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן":

מספר מכוניות למשפחה	1	2	3	4	5
שכיחות	65	150	220	140	55

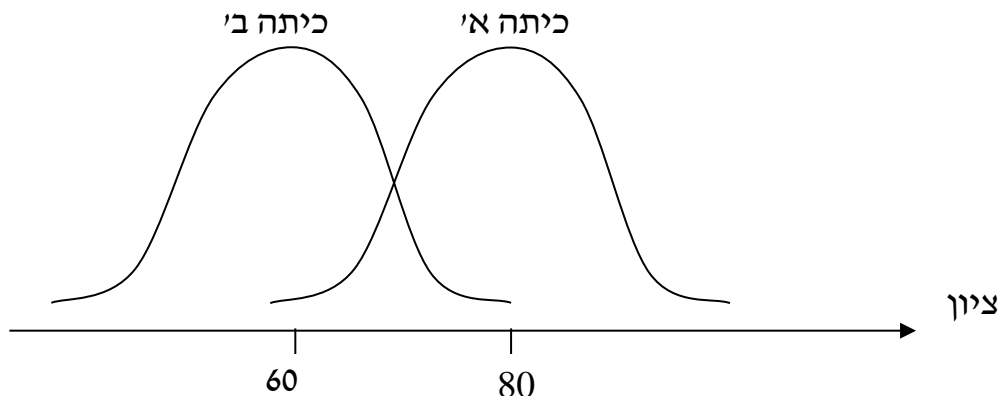
מהו הטווח הבין-רבעוני של מספר המכוניות למשפחה בישוב "הגורן"?

2) בסקר שנעשה בדקו את מספר ימי המחלה השנתיים של מורים בארץ.



- א. מה מייצגים הערכים בציר האופקי?
- ב. מהו הטווח הבין-רבעוני של מספר ימי המחלה של המורים?
- ג. אם נוסף 25 מורים אשר הצהירו שמספר ימי המחלה השנתיים שלהם הוא 4 ימים, כיצד הדבר ישנה את הטווח הבין-רבעוני? הסבירו.
- ד. אם מסתבר שחלק מהמורים בסקר הצהירו שהם חלו 7 ימים בשנה אבל בפועל הם חלו 8 ימים, כיצד הדבר ישנה את הטווח הבין-רבעוני? הסבירו.

3) לפניך שתי עקומות המתארות את התפלגות הציונים בכל כיתה. באיזו כיתה הטווח הבין-רבעוני גדול יותר?



- א. כיתה א.
- ב. כיתה ב'.
- ג. לשתיהן אותו טווח בין-רבעוני.
- ד. לא ניתן לדעת, אין מספיק נתונים.

4) הוספת גודל קבוע לכל תצפיות סדרת נתונים :

- א. תגדיל את הטווח הבין-רבעוני.
- ב. תקטין את הטווח הבין-רבעוני.
- ג. לא תשנה הטווח הבין-רבעוני.
- ד. לא ניתן לדעת מה יקרה לטווח הבין-רבעוני.

5) חושב הטווח הבין-רבעוני עבור התפלגות מסוימת והתקבלה התוצאה אפס. לכן :

- א. לפחות 50% מהתצפיות זהות.
- ב. סטיית התקן היא אפס.
- ג. ההתפלגות היא סימטרית.
- ד. מצב זה כלל לא יתכן.

- 6) סניף מספר 543 של בנק "רווה" בדק ל-80 לקוחות את מספר הפעמים שכל לקוח נכנס לסניף הבנק במשך שבוע. התוצאות שהתקבלו הן:
- 50 אנשים נכנסו 0 פעמים לסניף.
 - 20 אנשים נכנסו פעם אחת לסניף.
 - 5 אנשים נכנסו פעמיים לסניף.
 - 5 אנשים נכנסו יותר מפעמיים.
- מהו הטווח הבין-רבעוני?
- א. 60.
 - ב. 2.
 - ג. 50.
 - ד. 1.

- 7) התפלגות הציונים במבחן ווקסלר היא סימטרית לכן:
- א. טווח הציונים הוא אפס.
 - ב. הטווח הבין רבעוני של הציונים אפס.
 - ג. סעיפים א ו-ב הם נכונים.
 - ד. אף סעיף אינו נכון.

תשובות סופיות:

- 1) 2.
- 2) א. מספר ימי המחלה השנתיים. ב. 3. ג. יקטן. ד. לא ישתנה.
- 3) ג.
- 4) ג.
- 5) א.
- 6) ד.
- 7) ד.

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 6 - סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן

תוכן העניינים

1. כללי 17

סטטיסטיקה תיאורית – מדדי מיקום יחסי – ציון תקן:

רקע:

המטרה למדוד איך תצפית ממוקמת ביחס לשאר התצפיות בהתפלגות.

ציון תקן:

הנוסחה לציון תקן של תצפית היא: $Z = \frac{X - \bar{X}}{S}$.

- ציון התקן נותן כמה סטיות תקן סוטה התצפית מהממוצע. כלומר, ציון התקן מעיד על כמה סטיות תקן התצפית מעל או מתחת לממוצע:
- ציון תקן חיובי אומר שהתצפית מעל הממוצע.
 - ציון תקן שלילי אומר שהתצפית מתחת לממוצע.
 - ציון תקן אפס אומר שהתצפית בדיוק בממוצע.

דוגמה (פתרון בהקלטה):

במקום עבודה מסוים, ממוצע המשכורות הוא 8 אלף ₪, עם סטית תקן של אלפיים ₪. באותו מקום עבודה ההשכלה הממוצעת של העובדים הנה 14 שנים, עם סטית תקן של 1.5 שנים. ערן מרוויח במקום עבודה זה 11 אלף ₪ והשכלתו 16 שנים. מה ערן יותר, באופן יחסי, משכיל או משתכר?

שאלות:

- 1) תלמידי כיתה ח' ניגשו למבחן בלשון ולמבחן במתמטיקה. להלן התוצאות שהתקבלו:

המקצוע	ממוצע	סטיית תקן
לשון	74	12
מתמטיקה	80	16

עודד קיבל: 68 בלשון ו-70 במתמטיקה.

- א. באיזה מקצוע עודד טוב יותר באופן יחסי לשכבה שלו?
 ב. איזה ציון עודד צריך לקבל במתמטיקה כדי שיהיה שקול לציונו בלשון?
- 2) במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית (מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום. להלן טבלה המסכמת את המידע שנאסף על שני המשתנים:

ממוצע	תפוקה	מספר פועלים
ממוצע	48	15
סטיית תקן	10	2

- באחד הימים מתוך כלל הימים שנבדקו התפוקה הייתה 50 מאות מצברים ובאותו היום עבדו 13 פועלים. מה יותר חריג באותו היום, יחסית לשאר הימים שנבדקו: נתוני התפוקה או כמות הפועלים?
- א. התפוקה.
 ב. כמות הפועלים.
 ג. חריגים באותה מידה.
 ד. חסרים נתונים כדי לדעת זאת.

- 3) הגובה הממוצע של המתגייסים לצבא הוא 175 סנטימטר עם סטיית תקן של 10 סנטימטר. המשקל הממוצע הוא 66 ק"ג עם סטיית תקן של 8 ק"ג. ערך התגייס כשגובהו 180 ס"מ ומשקלו 59 ק"ג.
- א. במה ערך חריג יותר ביחס לשאר המתגייסים, גובהו או משקלו?
 ב. כמה ערך אמור לשקול כדי שמשקלו יהיה שקול לגובהו?

תשובות סופיות:

- 1) א. לשון. ב. 72.
 2) ב'.
 3) א. משקל. ב. 70.

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 7 - סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה

תוכן העניינים

1. כללי 19

סטטיסטיקה תיאורית – מדדי מיקום יחסי – אחוזונים בטבלה בדידה:

רקע:

האחוזון (המאון) ה- p הוא הערך בנתונים המחלק את הנתונים בצורה כזאת, שעד אליו (כולל) יש $p\%$ מהנתונים. מסמנים את האחוזון ה- p ב- X_p .

חישוב האחוזון מתוך נתונים בטבלת שכיחויות בדידה:

האחוזון הוא הערך שבו בפעם הראשונה השכיחות היחסית המצטברת (באחוזים) גדולה או שווה ל- $p\%$.

דוגמה (פתרון בהקלטה):

בסניף בנק 250 לקוחות. ספרו לכל לקוח את מספר תוכניות החיסכון שלו:

שכיחות יחסית מצטברת	שכיחות מצטברת	$F(x)$	# תוכניות החיסכון
		100	0
		75	1
		25	2
		25	3
		25	4

א. מצאו את האחוזון ה-25.

ב. מצאו את הערך ש-20% מהמקרים מעליו.

שאלות:

(1) להלן התפלגות של משתנה כלשהו:

$F(x)$	X
10	0
40	1
30	2
15	3
5	4

מצאו להתפלגות את:

- א. האחוזון ה-60.
- ב. המאון ה-40.
- ג. העשירון העליון.
- ד. הטווח בין הרבעונים.

(2) להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן":

מספר מכוניות למשפחה	1	2	3	4	5
שכיחות	65	150	220	140	55

חשבו את:

- א. העשירון התחתון.
- ב. האחוזון ה-30.
- ג. הערך ש-20% מהתצפית גדולות ממנו.
- ד. רבעון עליון.

תשובות סופיות:

- (1) א. 2 ב. 1 ג. 3 ד. 1
- (2) א. 1 ב. 2 ג. 4 ד. 4

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 8 - סטטיסטיקה תיאורית-מקדם ההשתנות

תוכן העניינים

1. כללי 21

סטטיסטיקה תיאורית – מקדם ההשתנות:

רקע:

כאשר מחשבים סטיית תקן למספר קבוצות בעלי ממוצע שונה, השוואת מידת פיזור הנתונים אינה מתייחסת לערך מרכז הנתונים (לממוצע למשל). על מנת לתת מדד פיזור המתחשב בממוצע הנתונים נחשב את מקדם ההשתנות –

$$CV = \frac{S(X)}{\bar{X}} : \text{Coefficient of Variation}$$

ככל שמקדם ההשתנות נמוך יותר, כך המשתנה מרוכז יותר סביב הממוצע, וככל שמקדם ההשתנות גבוה יותר, מידת הפיזור סביב הממוצע גבוהה יותר.

שאלות:

1) להלן נתונים לגבי ציונים במבחן באנגלית ב-3 כיתות מתוך שכבה י' בתיכון:

כיתה	ממוצע	מס' תלמידים	סטיית תקן
1	76	40	12
2	68	20	15
3	82	30	10

א. חשבו את מקדם ההשתנות בכל כיתה.

ב. מהי הכיתה הכי הטרוגנית?

2) נתונות שתי קבוצות: הממוצע בקבוצה א' הוא 100 והשונות 100.

הממוצע בקבוצה ב' הוא 500 והשונות 400.

באיזו קבוצה מידת הפיזור יחסית קטן יותר?

3) במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית

(מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום.

להלן טבלה המסכמת את האינפורמציה שנאספה על שני המשתנים:

מספר פועלים	תפוקה	ממוצע
15	48	
2	10	סטיית תקן

לפי קריטריון CV:

א. הפיזור באופן יחסי שווה בין התפוקה היומית לכמות הפועלים העובדים ביום.

ב. הפיזור יחסית יותר גדול עבור התפוקה היומית מאשר עבור מספר הפועלים ביום.

ג. הפיזור יחסית יותר גדול עבור מספר הפועלים ביום מאשר עבור התפוקה היומית.

ד. אין מספיק נתונים כדי לחשב את CV.

תשובות סופיות:

1) א. $\frac{\sigma}{\bar{X}}$. ב. כיתה ב'.

2) קבוצה ב'.

3) ב'.

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 9 - סטטיסטיקה תיאורית שאלות אמריקאיות

תוכן העניינים

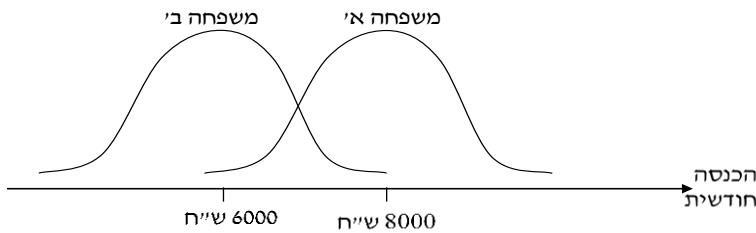
1. כללי 23

סטטיסטיקה תיאורית – שאלות אמריקאיות:

שאלות:

שאלות 1-3 מתייחסות לקטע הבא:

להלן שתי עקומות המתארות את התפלגות ההכנסות החודשיות של שתי משפחות שנבחרו באקראי:



- (1) לאיזו משפחה הכנסה שכיחה גבוהה יותר?
- משפחה א'.
 - משפחה ב'.
 - לשתיהן אותה הכנסה שכיחה.
 - לא ניתן לדעת – אין מספיק נתונים.
- (2) באיזו משפחה ההכנסה החציונית שווה להכנסה הממוצעת?
- משפחה א'.
 - משפחה ב'.
 - בשתיהן ההכנסה החציונית שווה להכנסה הממוצעת.
 - לא ניתן לדעת – אין מספיק נתונים.
- (3) באיזו משפחה סטית התקן של ההכנסה החודשית גבוהה יותר?
- משפחה א'.
 - משפחה ב'.
 - לשתיהן אותה סטית תקן.
 - לא ניתן לדעת – אין מספיק נתונים.

הנתונים הבאים מתייחסים לשאלות 4-6:

להלן נתונים חלקיים של טבלת שכיחויות:
כמו כן, נתון כי הממוצע הוא 1.66.

$F(x)$	x
?	0
10	1
6	2
15	3
?	4
50	סה"כ

(4) השכיח של הנתונים הוא:

- א. 0.
- ב. 15.
- ג. ישנם שני שכיחים: 0 ו-3.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של השכיח.

(5) חציון הנתונים הוא:

- א. 2.
- ב. 1.5.
- ג. 25.5.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של החציון.

(6) הטווח של הנתונים:

- א. 11.
- ב. 3.
- ג. 4.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של החציון.

(7) בהתפלגות אסימטרית ימנית של משתנה כמותי רציף, הערך המתאים למאון ה-30, ציון התקן שלו הוא בהכרח:

- א. שלילי.
- ב. חיובי.
- ג. אפס.
- ד. לא ניתן לדעת ללא הנתונים.

- 8) סדרת נתונים סטטיסטיים מונה 10 תצפיות. נתון כי סדרת הנתונים סימטרית סביב הממוצע. ממוצע הסדרה-40 ושונות הסדרה-100. בשלב מאוחר יותר נוספו שתי תצפיות נוספות לסדרה : 50 ו-30. השונות של 12 התצפיות :
- א. תקטן.
 - ב. תגדל.
 - ג. לא תשתנה.
 - ד. לא ניתן לחשב את השונות ללא ידיעת התצפיות.

הנתונים הבאים מתייחסים לשאלות 9-10:

בחברת "טיק" המשכורת הממוצעת היא 4,600 ₪ וסטיית התקן של משכורת זו הינה 200 ₪. לאחר מו"מ עם ועד עובדי ההנהלה סוכם כי המשכורת תוכפל פי 1.5.

- 9) מהי המשכורת הממוצעת החדשה (ב-₪)?

- א. 2,300.
- ב. 6,900.
- ג. 4,650.
- ד. 4,600.

ה. חסרים נתונים כדי לדעת.

- 10) מהי סטיית התקן של המשכורת לאחר יישום המו"מ לגבי השכר (ב-₪)?

- א. 200.
- ב. 300.
- ג. 675.
- ד. לא ניתן לדעת.

- 11) הוספת גודל קבוע לכל תצפיות סדרת נתונים :

- א. תגדיל את סטיית התקן.
- ב. תקטין את סטיית התקן.
- ג. לא תשנה את סטיית התקן.
- ד. לא ניתן לדעת.

הנתונים הבאים מתייחסים לשאלות 12-13:

להלן נתונים על ציוני תלמידים שנבחנו במועדים שונים בסטטיסטיקה:

שם התלמיד	ציון	ממוצע הציונים במועד בו נבחן	סטיית התקן של הציונים במועד בו נבחן
צבי	50	50	12
סטף	82	80	5
שרית	65	60	15
לובה	60	63	1.5
מיטב	70	70	10

12 התלמיד הטוב ביותר ביחס לנבחנים באותו מועד בו נבחן הוא:

- א. מיטב.
- ב. צבי.
- ג. לובה.
- ד. שרית.
- ה. סטף.

13 פנינה נבחנה עם סטף וציון התקן שלה שווה לציון התקן של שרית לכן ציונה הוא:

- א. 80.55.
- ב. 65.
- ג. 80.
- ד. 81.66.

הנתונים הבאים מתייחסים לשאלות 14-17:

בבדיקת פתע של משרד הבריאות במפעל שוקולד, נמצא ש:

שוקולד פגום	0	1	2	3	4	5	6	7
מס' קופסאות	35	63	48	12	13	11	10	8

14 מהו החציון של מספר הפגומים בקופסא:

- א. 1.
- ב. 2.
- ג. 4.
- ד. לא ניתן לדעת.

15 מהו הרבעון התחתון של מספר הפגומים בקופסא?

- א. 1.
- ב. 2.
- ג. 3.
- ד. 4.
- ה. לא ניתן לדעת.

16) מספר הפגומים בקופסא הוא משתנה :

- א. סדר.
- ב. שמי.
- ג. כמותי בדיד.
- ד. כמותי רציף.

17) השכיח של מספר הפגומים בקופסא :

- א. 63.
- ב. 1.
- ג. 200.
- ד. לא ניתן לדעת.

18) ביחס לציר המספרים, רוב הערכים בהתפלגות א-סימטרית ימנית נמצאים :

- א. בערכים הגבוהים.
- ב. בחלוקה זהה בין הערכים הגבוהים והנמוכים.
- ג. בערכים הנמוכים.
- ד. לא ניתן לדעת.
- ה. אף לא תשובה מהני"ל נכונה.

19) בוצע מחקר על מספר העובדים בחברות מזון לעומת חברות תקשורת. החציון והממוצע בשתיהן שווה 8.

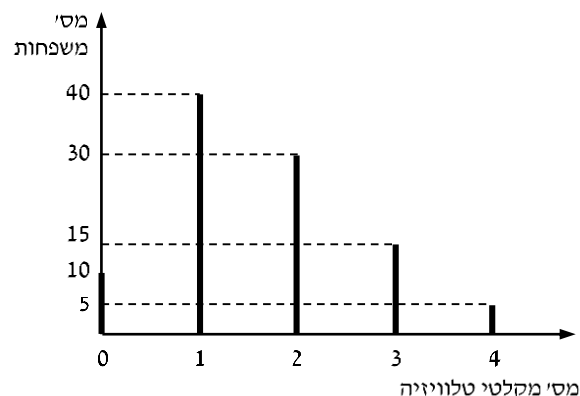
איזה מהטענות הבאות היא הנכונה והמלאה ביותר :

- א. השכיחות ב-2 החברות זהה אך שונה מ-8.
- ב. השכיח ב-2 החברות זהה אך לא ניתן לדעת מהו.
- ג. השכיח בשתי חברות הינו בהכרח 8.
- ד. שכיח בחברה אחת שונה מ-8 ובשנייה הוא 8.
- ה. אף תשובה אינה נכונה.

הנתונים הבאים מתייחסים לשאלות 20 עד 24 :

נערך סקר על מספר מקלטי הטלוויזיה הנמצאים בבית.

תוצאות הסקר נתונות בדיאגרמת מקלות הבאה :



(20) המשתנה הנחקר כאן הוא :

- א. משתנה שמי.
- ב. משתנה מסולם סדר.
- ג. משתנה כמותי בדיד.
- ד. משתנה כמותי רציף.

(21) הטווח של ההתפלגות הוא :

- א. 35.
- ב. 4.
- ג. 3.
- ד. 2.

(22) ממוצע מספר מקלטי הטלוויזיה למשפחה הוא :

- א. 1.65.
- ב. 1.5.
- ג. 1.
- ד. 2.

(23) השכיח של התפלגות זו היא :

- א. 40.
- ב. 1.5.
- ג. 1.
- ד. 2.

(24) מסתבר שיש בין 2 ל-5 משפחות נוספות שאין להם מקלטי טלוויזיה ויש לצרף את המשפחות הללו להתפלגות. כיצד הנתון זה ישפיע על סטיית התקן?

- א. יקטין אותו.
- ב. יגדיל אותו.
- ג. לא ישנה אותו.
- ד. אין לדעת.

תשובות סופיות:

(5 ב'	(4 ג'	(3 ג'	(2 ג'	(1 א'
(10 ב'	(9 ב'	(8 ג'	(7 א'	(6 ג'
(15 א'	(14 ב'	(13 ד'	(12 ה'	(11 ג'
(20 ג'	(19 ה'	(18 ג'	(17 ב'	(16 ג'
	(24 ב'	(23 ג'	(22 א'	(21 ב'

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 10 - מקדם המתאם (מדד קשר) הלינארי ומובהקותו

תוכן העניינים

1. מקדם המתאם הלינארי (פירסון)..... 30
2. חישוב מקדם המתאם של פירסון (לנארי) במדגם (ללא ספר)

מקדם המתאם (מדד קשר) הלינארי ומובהקותו

מדד הקשר הלינארי (פירסון) – מבוא

מעוניינים לבדוק עד כמה קיים קשר מסוג קשר לינארי (קו ישר) בין שני משתנים. שני המשתנים שאנו בודקים לגביהם קשר צריכים להיות משתנים כמותיים. מבחינת סולמות מדידה כל משתנה נחקר צריך להיות מסולם רווחים או מנה. בדרך כלל המשתנה המוצג כ- Y הוא המשתנה התלוי והמשתנה המוצג ב- X הוא המשתנה הבלתי תלוי. תיאור גרפי לנתונים נעשה על ידי דיאגרמת פיזור. בדיאגרמת פיזור אנחנו מסמנים כל תצפית בנקודה לפי שיעור ה- X ושיעור ה- Y שלה. דיאגרמת הפיזור נותנת אינדיקציה גרפית על הקשר בין שני המשתנים.

דוגמה (פתרון בהקלטה) :

בבניין 8 דירות בדקו לכל דירה את מספר החדרים שלה וכמו כן את מספר הנפשות הגרות בדירה. להלן התוצאות שהתקבלו :

4	4	3	3	2	3	2	2	מספר חדרים בדירה
5	4	4	3	2	2	1	0	מספר הנפשות בדירה

- (1) כמה תצפיות ישנן בדוגמה?
- (2) כמה משתנים ישנם בדוגמה, מי הם?
- (3) שרטטו לנתונים דיאגרמת פיזור.
- (4) מי המשתנה התלוי ומיהו המשתנה הבלתי תלוי?

מקדם מתאם 1-או 1 אומר שקיים קשר לינארי מלא בין המשתנים שניתן לבטאו על ידי נוסחה של קו ישר: $y = ax + b$.

מתאם חיובי מלא (מקדם מתאם 1):

קיים קשר לינארי מלא בו השיפוע a יהיה חיובי ואילו מתאם שלילי (מקדם מתאם-1) מלא אומר שקיים קשר לינארי מלא בו השיפוע a שלילי.

מתאם חיובי חלקי:

ככל שמשנתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל שמשנתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט. ככל שמקדם המתאם קרוב לאפס עוצמת הקשר יותר חלשה וככל שהמדד רחוק יותר מהאפס העוצמה יותר חזקה. לסיכום, מקדם המתאם בודק את עוצמת הקשר הלינארי, ואת כיוון הקשר.

מקדם המתאם הלינארי אינו מושפע מיחידות המדידה. כל שינוי ביחידות המדידה של המשתנים, לא ישנה את מקדם המתאם.

מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב: ρ - פרמטר המאפיין את עוצמת הקשר הלינארי באוכלוסייה וכיוונו בין שני המשתנים הנחקרים. כאשר:

r - מדד הקשר הלינארי במדגם שמהווה אומדן לפרמטר ρ .

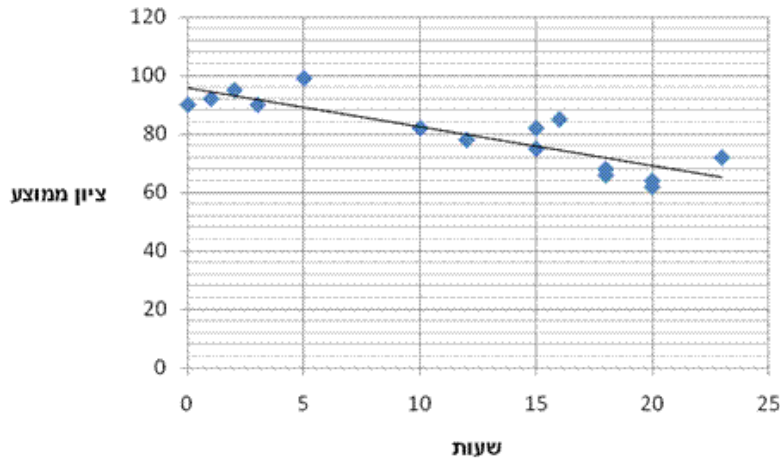
קיומו של מתאם בין שני משתנים אינו מצביע על סיבתיות בהכרח. למשל, אם נמצא מתאם חיובי בין כמות הסוכרזית שאדם אוכל לבין במשקל שלו אין זה אומר שהסיבה להשמנה היא הסוכרזית. מדד הקשר של פירסון הוא מדד קשר סימטרי, כלומר אם נחליף את X ב- Y התוצאה תהיה זהה.

דוגמה (פתרון בהקלטה):

- מה ניתן להגיד על מקדם המתאם של שני המשתנים על סמך דיאגרמת הפיזור ששרטטנו?
- אם היינו משנים את השרטוט כך שבציר האנכי היה המשתנה "מספר החדרים" ובציר האופקי היה "מספר הנפשות", האם הדבר היה משפיע על מדד הקשר של פירסון?

שאלות

1) חוקר רצה לאפיין את הקשר בין מספר השעות בשבוע שסטודנט מקדיש לבילויים לבין הציון הממוצע שלו בסוף הסמסטר. לשם כך הוא אסף נתונים של 15 סטודנטים ויצר דיאגרמת פיזור:



- א. מיהו המשתנה הבלתי תלוי?
- ב. מה ניתן לומר על כיוון הקשר בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר? מה ניתן להגיד על עוצמת הקשר?

2) להלן טבלה המסכמת את מקדמי המתאם הלינארי בין ציוני מבחנים שונים שהתקבלו עבור תלמידים בכיתה מסוימת:

מתמטיקה	לשון	ספורט	
?	-0.7	?	ספורט
0.6	?	?	לשון
?	?	-0.1	מתמטיקה

- א. השלימו את מקדמי המתאם שמסומנים בסימן שאלה בטבלה.
- ב. בין אילו שני ציוני מקצועות שונים קיים מתאם בעל העוצמה החזקה ביותר?

3) במחקר נתבקשו לבדוק את הקשר בין מספר שעות התרגול של קורס לבין הציון הסופי שלו. להלן תוצאות מדגם שהתקבל:

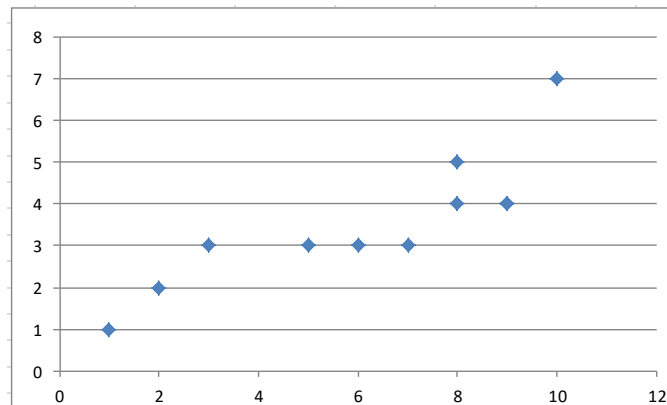
שעות תרגול	ציון סופי
20	90
25	90
30	95
15	60
30	90
20	85
10	50

- א. מיהו המשתנה התלוי ומיהו המשתנה הבלתי תלוי בדוגמה זו?
- ב. שרטטו דיאגרמת פיזור לנתונים.
- ג. מה ניתן לומר על הקשר בין המשתנים במדגם?
- ד. מסתבר שבסופו של דבר נתנו פקטור של 5 נקודות לציון הסופי. כיצד הדבר היה משנה את מקדם המתאם של המדגם?

4) בתחנה המטאורולוגית רצו לבדוק את הקשר שבין הטמפרטורה במעלות צלזיוס לכמות המשקעים במ"מ. הם אספו נתונים על 10 ימים במהלך חודש ינואר. המתאם שהתקבל היה 0.8.

- א. השלימו את המשפט:
בחודש ינואר ככל שהטמפרטורה היומית נוטה לרדת, כך כמות המשקעים נוטה _____.
- ב. הוחלט להעביר את הטמפרטורה למעלות פרנהייט על מנת שיוכלו להשוות אותה לנתונים מארה"ב. נוסחת המעבר היא $F^0 = 32 + \frac{9}{5}C^0$.
כיצד הדבר ישפיע על מקדם המתאם בין הטמפרטורה במעלות פרנהייט לכמות המשקעים במ"מ?

5) להלן דיאגרמת פיזור המראה קשר בין שני משנים:

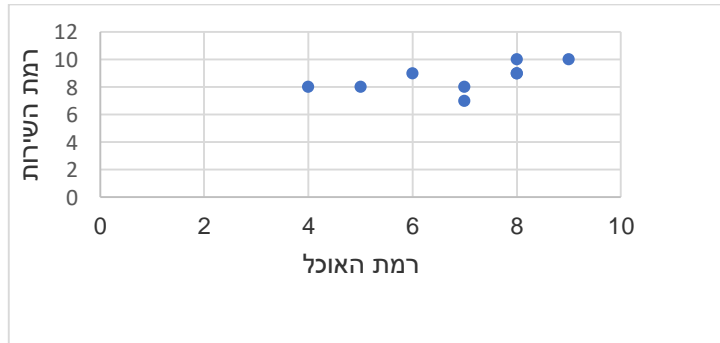


- א. השלימו: ניתן לראות שהקשר הוא לינארי _____ (מלאו חלקי) כיוון הקשר הוא (חיובי/שלילי).
- ב. השלימו: אם היינו מוסיפים תצפית שערך ה- X שלה הוא 4 וערך ה- Y שלה הוא 7, מקדם המתאם של פירסון היה _____ (גדלו קטן/לא משתנה).

שאלות רב ברירה (יש לבחור את התשובה הנכונה):

- 6) חוקר אקלים דגם כמה ימים בשנה ומדד את הטמפרטורה בטורונטו שבקנדה ואת הטמפרטורה בסידני שבאוסטרליה באותו היום. הוא חישב ומצא מקדם מתאם שלילי בין הטמפרטורה היומית בטורונטו לבין הטמפרטורה היומית בסידני. משמעות מקדם המתאם השלילי במדגם:
- א. אין קשר בין הטמפרטורה בטורונטו לבין הטמפרטורה בסידני בימים שנדגמו.
ב. במדגם, רוב הטמפרטורות בטורונטו היו שליליות.
ג. ההפרש בין הטמפרטורה בטורונטו לבין הטמפרטורה באוסטרליה, במדגם זה, הוא שלילי.
ד. במדגם יש נטייה שהטמפרטורה יורדת בטורונטו לטמפרטורה לעלות בסידני.

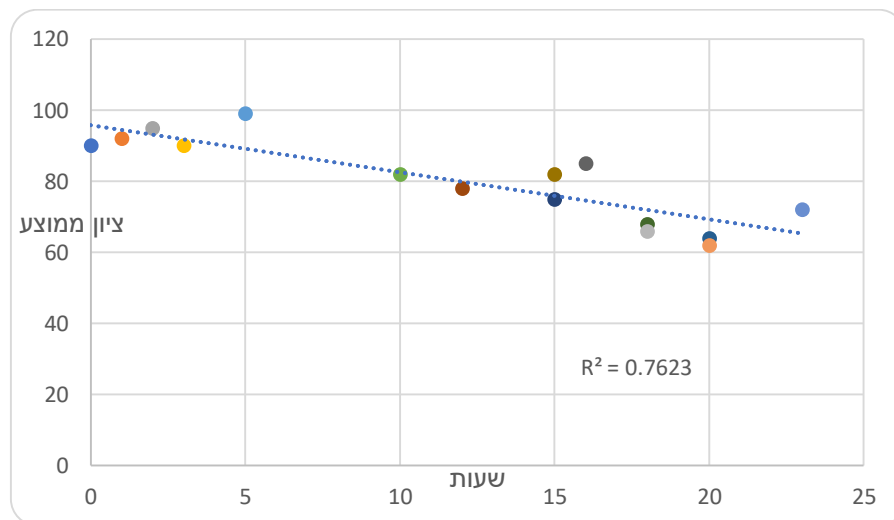
- 7) בסקר שביעות רצון שנערך בבית הקפה "פת לחם" התבקשו הלקוחות לדרג את מידת שביעות הרצון שלהם (בסולם 1-10) בשני נושאים: רמת האוכל ורמת השירות.



מה יהיה ערכו של מקדם המתאם (r)?

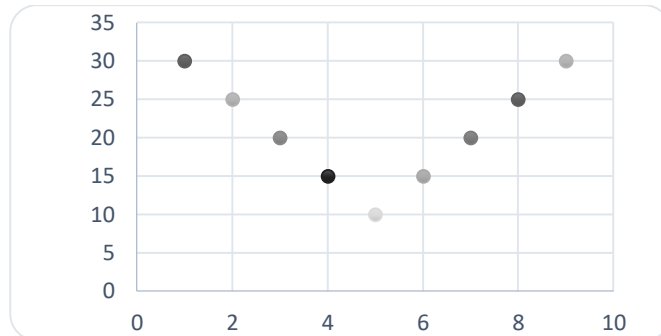
- א. $r = -0.3$
 ב. $r = 0$
 ג. $r = 1.125$
 ד. $r = 0.593$

- 8) חוקר רצה לאפיין את הקשר בין מספר השעות בשבוע שסטודנט מקדיש לבילויים לבין הציון הממוצע שלו בסוף הסמסטר. לשם כך הוא אסף נתונים של 15 סטודנטים ויצר דיאגרמת פיזור.



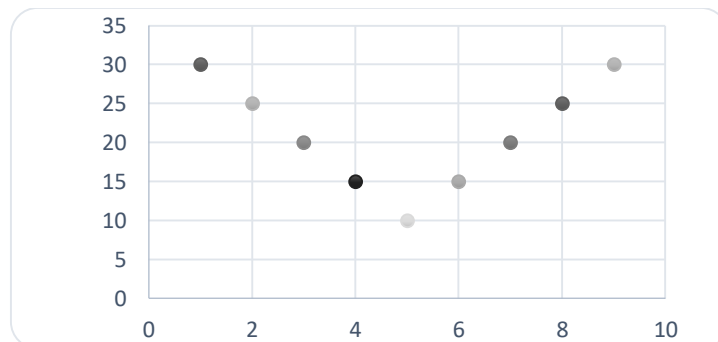
- מה ניתן לומר על כיוון הקשר במדגם בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר?
- א. ככל שמבלים יותר הציון נוטה לרדת.
 ב. אין קשר בין שעות הבילוי לציון.
 ג. ככל שמבלים פחות הציון נוטה לרדת.
 ד. ככל שהציון נוטה לרדת הסטודנט מבלה פחות.

9) התרשים הבא מתאר קשר בין שני משתנים, איזה מהמתאמים הבאים הוא המתאים ביותר לתיאור הקשר בין שני המשתנים?



- א. $r = 1$ היות ושני המשתנים יוצרים קוים ישרים.
 ב. $r = 2$ היות ויש שני קוים בעלי קשר מושלם.
 ג. $r = 0$ היות והקו יורד ואחר כך עולה באותו האופן.
 ד. $r = \pm 1$ היות ויש קו עולה וגם קו יורד.

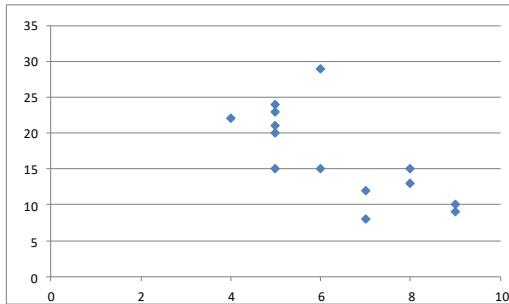
10) התרשים הבא מתאר דיאגרמת פיזור.



איזו טענה נכונה?

- א. בתרשים מוצג הקשר בין שני משתנים.
 ב. בתרשים מוצג הקשר בין 9 משתנים.
 ג. בתרשים מוצג הקשר בין 10 משתנים.
 ד. אין לדעת כמה משתנים מוצגים בתרשים.

בגרף הבא מתוארת דיאגרמת פיזור של שני משתנים:



X - (משתנה בלתי תלוי בציר האופקי)

Y - (משתנה תלוי).

במדגם התקבל $r^2 = 0.52$.

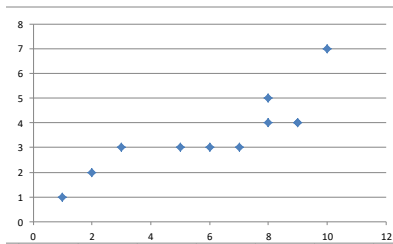
11) לאור הנתונים המופיעים בדיאגרמה, איזה מבין הערכים הבאים מתאים להיות התוצאה של r ?

- א. -0.52
- ב. 0.72
- ג. -0.72
- ד. 0.52

12) אם מקדם המתאם בין שני משתנים הוא 1, אזי:

- א. הערכים של המשתנים הם חיוביים.
- ב. עבור כל תצפית ערך של משתנה אחד שווה לערך של המשתנה השני.
- ג. הקשר הלינארי הוא בעוצמה חזקה.
- ד. אף אחת מהתשובות לא בהכרח נכונה.

13) להלן דיאגרמת פיזור:



מה יהיה מקדם המתאם בין שני המשתנים?

- א. 1
- ב. 0.85
- ג. 0.15
- ד. 0

14) בבדיקת קשר בין שני משתנים התקבל: $r = -1$.

- א. קיימת נוסחה לינארית הקושרת בין כל התצפיות.
- ב. לא קיים קשר בין שני המשתנים.
- ג. ככל שמשתנה אחד נוטה לרדת גם לשני יש נטייה לרדת.
- ד. קיים קשר בין שני המשתנים, אך לא ניתן לדעת מאיזה סוג.

15) לפי הפתגם "רחוק מהעין, רחוק מהלב", יש קשר _____ בין קרבה פיזית לקרבה נפשית.

- א. חיובי
- ב. שלילי
- ג. אפסי
- ד. לא ניתן לדעת.

16) מבחן אמי"ר הינו מבחן מיון באנגלית של המרכז הארצי לבחינות והערכה. הציון המינימלי בבחינה הינו 150 והמקסימלי הינו 250. בקורס הכנה למבחן השתתפו 19 תלמידים. להלן הציונים שלהם על פי פלט שהתקבל:

	159
	170
	180
	185
	204
	224
	236
	212
	168
	189
	195
	163
	187
	206
	201
	223
	242
	203
	205
197.47	AVERAGE
536.25	VARPA

יש להוסיף עמודה נוספת לצד עמודת הציונים שתראה לכל תלמיד כמה נקודות חסרות לו כדי להשלים לציון המקסימלי בבחינה.

מה יהיה מקדם המתאם בין שתי העמודות (כלומר, מקדם המתאם בין הציון לבין הנקודות החסרות)?

- א. -1
- ב. 1
- ג. -0.5
- ד. 0.5

17) מקדם המתאם בין שטחי דירה למחיר שלהם חושב ונמצא 1.2. מה נובע מכך?

- א. ככל שהדירה גדולה יותר בשטחה כך היא יקרה יותר.
- ב. ככל שהדירה קטנה יותר בשטחה כך היא זולה יותר.
- ג. לא קיים קשר בין שטח הדירה למחיר הדירה.
- ד. מצב כזה שמתואר הנתונים לא אפשרי.

18) אם ניקח 10 אנשים ונרשום לכל אדם את הגובה במטר וכמו כן את הגובה בס"מ. מה יהיה מקדם המתאם בין גובה האדם במטר לגובה האדם בס"מ?

- א. 1
- ב. 0
- ג. -1
- ד. לא ניתן לדעת.

- 19) נמצא מתאם חיובי בעוצמה גבוהה בין X – ציון בבגרות בלשון ל Y – ציון בבגרות במתמטיקה. אילו מהמשפטים הבאים נכון?
- א. ניתן לומר שאחת מהסיבות להבדלים שיש לסטודנטים במתמטיקה נובעים מההבדלים שיש להם בלשון.
- ב. קיימת נוסחה של קו ישר שקושרת בין ציון בבגרות במתמטיקה לציון בבגרות בלשון.
- ג. ללא יוצא מן הכלל, ניתן להגיד שכל תלמיד שמצליח יותר מתלמיד אחר בלשון גם יצליח יותר מאותו תלמיד במתמטיקה.
- ד. אף אחד מהטענות שהוצגו אינה בהכרח נכונה.

- 20) עבור סדרה של תצפיות מדדו את X ואת Y . נמצא שעבור כל התצפיות שהערך של Y ירד הערך של X בהכרח ירד ללא יוצא מן הכלל. מקדם המתאם של פירסון יהיה בהכרח:
- א. 1
- ב. -1
- ג. 0
- ד. אף אחת מהתשובות.

תשובות סופיות

- (1) א. שעות בילוי.
 ב. הקשר חלקי, כיוון הקשר שלילי.
 (2) א. להלן טבלה:

מתמטיקה	לשון	ספורט	
0.1	-0.7	1	ספורט
0.6	1	-0.7	לשון
1	0.6	-0.1	מתמטיקה

- (3) א. ב"ת- מס' שעות התרגול, תלוי- ציון.
 ג. קשר לינארי חיובי חלקי.
 (4) א. לעלות.
 (5) א. חלקי, חיובי.
 ב. ראה גרף בפתרון וידאו.
 ד. מקדם המתאם לא היה משתנה.
 ב. לא ישפיע על מקדם המתאם.
 ב. קטן.

- (6) ד' (7) ד' (8) א' (9) ג' (10) א'
 (11) ג' (12) ד' (13) ב' (14) א' (15) א'
 (16) א' (17) ד' (18) א' (19) ד' (20) ד'

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 11 - מדד הקשר של ספירמן

תוכן העניינים

1. הקדמה (ללא ספר)

2. חישוב מדד הקשר של ספירמן (ללא ספר)

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 12 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

תוכן העניינים

1. התפלגות נורמלית (טבלת z כוללת ערכים שליליים) (ללא ספר)

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 13 - הסקה סטטיסטית - הקדמה

תוכן העניינים

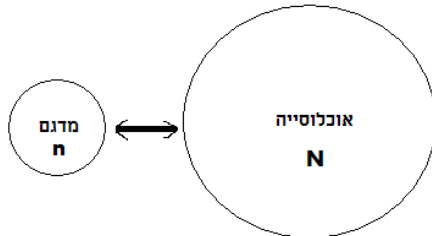
1. כללי 41

הסקה סטטיסטית – הקדמה:

רקע:

אוכלוסייה:

קבוצה שאליה מפנים שאלה מחקרית. למשל, חברת תרופות שמעוניינת לפתח תרופה למחלת הסוכרת מתעניינת באוכלוסיית חולי הסוכרת בעולם.



מדגם:

חלק מתוך האוכלוסייה. למשל, אם נדגום באקראי 10 אנשים מתוך חולי הסוכרת אז זהו מדגם מתוך אוכלוסיית חולי הסוכרת.

במקרים רבים אין אפשרות לחקור את כל האוכלוסייה כיוון שאין גישה לכולה, היא גדולה מידי, או מוגבלים בזמן ובאמצעים טכניים ולכן מבצעים מדגם במטרה לבצע הסקה סטטיסטית מהמדגם לאוכלוסייה. הדגימה בקורס תהיה דגימה מקרית - הכוונה לדגימה שבה לכל תצפית באוכלוסייה יש את אותו סיכוי להיכלל במדגם.

סטטיסטי:

גודל המחושב על המדגם.

פרמטר:

גודל המתאר את האוכלוסייה.

הסימונים לפרמטר וסטטיסטי הם שונים:

פרמטר (אוכלוסייה)	סטטיסטי (מדגם)	ממוצע
μ	\bar{X}	
P	\hat{p}	פרופורציה (שכיחות יחסית)

פרמטר הוא גודל קבוע גם אם אנו לא יודעים אותו סטטיסטי הוא משתנה ממדגם למדגם ולכן יש לו התפלגות הנקראת התפלגות הדגימה.

דוגמה (פתרון בהקלטה):

25% מאזרחי המדינה תומכים בהצעת החוק של חבר כנסת מסוים. הוחלט לדגום 200 אזרחים ומתוכם לבדוק מהו אחוז התומכים בהצעת החוק.

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מה הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו הסטטיסטי שמתכננים להוציא מהמדגם?

ו. האם הפרמטר או הסטטיסטי הוא משתנה מקרי?

שאלות:

- (1) מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
- מי האוכלוסייה?
 - מה המשתנה?
 - מהם הפרמטרים?
 - מהו גודל המדגם?
- (2) להלן התפלגות מספר מקלטי הטלוויזיה למשפחה בישוב "העוגן". נגדיר את X להיות מספר המקלטים של משפחה אקראית. מתכננים לדגום מאוכלוסייה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלוויזיה במדגם.
- מיהי האוכלוסייה ומהו המשתנה הנחקר?
 - מהו הסטטיסטי שיילקח מהמדגם ומה סימונו?

מספר משפחות	מספר מקלטים
50	0
250	1
350	2
300	3
50	4
סך הכול $N = 1000$	

- (3) נתון כי 20% מהשכירים במדינה הם אקדמאיים. נבחרו באקראי 10 שכירים באותה אוכלוסייה ומתכננים לפרסם את מספר האקדמאיים שנדגמו.
- מיהי האוכלוסייה?
 - מה המשתנה באוכלוסייה?
 - מהם הפרמטרים?
 - מהו הסטטיסטי?

תשובות סופיות:

- (1) א. כלל הסטודנטים במכללה שסיימו סטטיסטיקה א. ב. ציון. ג. ממוצע: 78, סטיית תקן: 15. ד. 2.
- (2) א. האוכלוסייה: 1000 משפחות בישוב העוגן, המשתנה הנחקר: מס' מקלטים. ב. \bar{X} = ממוצע מדגם.
- (3) א. השכירים במדינה. ב. השכלה: אקדמאי, לא אקדמאי. ג. שיעור ההצלחות באוכלוסייה: 0.2. ד. מס' האקדמאים במדגם.

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 14 - התפלגות הדגימה ומשפט הגבול המרכזי

תוכן העניינים

1. התפלגות ממוצע המדגם ומשפט הגבול המרכזי 44

התפלגות ממוצע המדגם ומשפט הגבול המרכזי:

רקע:

בפרק זה נדון בהתפלגות של ממוצע המדגם: $\bar{x} = \frac{\sum x_i}{n}$.

מכיוון שממדגם למדגם אנו יכולים לקבל ממוצע מדגם שונה, אזי ממוצע המדגם הוא משתנה מקרי ויש לו התפלגות.

גדלים המתארים התפלגות כלשהי או אוכלוסייה כלשהי נקראים פרמטרים.

להלן רשימה של פרמטרים החשובים לפרק זה:

ממוצע האוכלוסייה נסמן ב- μ (נקרא גם תוחלת).

שונות אוכלוסייה נסמן ב- σ^2 .

סטיית תקן של אוכלוסייה: σ .

תכונות התפלגות:

ממוצע כל ממוצעי המדגם האפשריים שווה לממוצע האוכלוסייה: $E(\bar{x}) = \mu_x = \mu$.

שונות כל ממוצעי המדגם האפשריים שווה לשונות האוכלוסייה מחולק ב- n .

תכונה זו נכונה רק במדגם מקרי: $V(\bar{x}) = \sigma_x^2 = \frac{\sigma^2}{n}$.

יש יחס הפוך בין גודל המדגם לבין שונות ממוצעי המדגם.

אם נוציא שורש לשונות נקבל סטיית תקן של ממוצע המדגם שנקראת גם

טעות תקן: $\sigma(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$.

דוגמה (פתרון בהקלטה):

השכר הממוצע במשק הינו 9000 ₪ עם סטיית תקן של 4000. דגמו באקראי 25 עובדים.

א. מיהי אוכלוסיית המחקר? מהו המשתנה הנחקר?

ב. מהם הפרמטרים של האוכלוסייה?

ג. מה התוחלת ומהי סטיית התקן של ממוצע המדגם?

דגימה מהתפלגות נורמאלית:

אם נדגום מתוך אוכלוסייה שהמשתנה בה מתפלג נורמאלית עם ממוצע μ ושוונות σ^2 .

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right), Z_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

ממוצע המדגם גם יתפלג נורמאלית:

דוגמה (פתרון בהקלטה):

משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.
 מה ההסתברות שבמדגם של 4 תינוקות אקראיים בעת הולדתם המשקל הממוצע של התינוקות יהיה מתחת ל-3.5 ק"ג?

משפט הגבול המרכזי:

אם אוכלוסייה מתפלגת כלשהו עם ממוצע μ ושוונות σ^2 אזי עבור מדגם מספיק

$$\bar{x} \rightsquigarrow N\left(\mu, \frac{\sigma^2}{n}\right) \text{ גדול } (n \geq 30) \text{ ממוצע המדגם מתפלג בקירוב נורמאלי:}$$

דוגמה (פתרון בהקלטה):

משקל חפיסת שוקולד בקו ייצור מתפלג עם ממוצע 100 גרם וסטיית תקן של 4 גרם.
 דגמו מקו הייצור 36 חפיסות שוקולד אקראיות.
 מה ההסתברות שהמשקל הממוצע של חפיסות השוקולד שנדגמו יהיה מתחת ל-102 גרם?

שאלות:

- (1) מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
- מיהי האוכלוסייה?
 - מה המשתנה?
 - מהם הפרמטרים?
 - מהו גודל המדגם?
 - מהו תוחלת ממוצע המדגם?
 - מהי טעות התקן?
- (2) משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.
- מה ההסתברות שתינוק אקראי בעת הלידה ישקול פחות מ-3800 גרם? נתון כי ביום מסוים נולדו 4 תינוקות.
 - מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק"ג?
 - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק"ג?
 - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?
 - הסבירו ללא חישוב כיצד התשובה לסעיף הקודם הייתה משתנה אם היה מדובר על יותר מ-4 תינוקות?
- (3) הגובה של המתגייסים לצה"ל מתפלג נורמאלית עם תוחלת של 175 ס"מ וסטיית תקן של 10 ס"מ. ביום מסוים התגייסו 16 חיילים.
- מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס"מ?
 - מה ההסתברות שהגובה הממוצע שלהם יהיה בדיוק 180 ס"מ?
 - מה ההסתברות שהגובה הממוצע שלהם יסטה מתוחלת הגבהים בפחות מ-5 ס"מ?
 - מהו הגובה שבהסתברות של 90% הגובה הממוצע של המדגם יהיה נמוך ממנו?

- (4) הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועות. האדם נוסע לעבודה במשך שבוע 5 פעמים. לצורך הפתרון הניחו שזמן הנסיעה לעבודה מתפלג נורמאלית.
- א. מה ההסתברות שבמשך שבוע משך הנסיעה הממוצע יהיה מעל 33 דקות?
 ב. מהו הזמן שבהסתברות של 90% ממוצע משך הנסיעה השבועי יהיה גבוה ממנו?
 ג. מה ההסתברות שממוצע משך הנסיעה השבועי יהיה מרוחק מ-30 דקות בלפחות 2 דקות?
 ד. כיצד התשובה לסעיף הקודם הייתה משתנה אם האדם היה נוסע לעבודה 6 פעמים בשבוע?
- (5) נפח היין בבקבוק מתפלג נורמאלית עם תוחלת של 750 סמ"ק וסטיית תקן של 10 סמ"ק.
- א. בארגו 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגו יהיה בדיוק 755 סמ"ק?
 ב. בארגו 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגו יהיה יותר מ-755 סמ"ק?
 ג. בארגו 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגו יהיה לפחות 755 סמ"ק?
 ד. בקבוקי היין שבארגו נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?
- (6) משתנה מתפלג נורמאלית עם תוחלת 80 וסטיית תקן 4.
- א. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה כאשר גודל המדגם הוא 9?
 ב. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה שגודל המדגם הוא 16?
 הסבר את ההבדל בתשובות של שני הסעיפים.
- (7) לפי הערכות הלשכה המרכזית לסטטיסטיקה השכר הממוצע במשק הוא 8000 ₪ עם סטיית תקן של 3000 ₪. מה ההסתברות שבמדגם מקרי של 100 עובדים השכר הממוצע יהיה יותר מ-8500 ₪?

8) אורך צינור שמפעל מייצר הינו עם ממוצע של 70 ס"מ וסטיית תקן של 10 ס"מ.

א. נלקחו באקראי 100 מוטות, מה ההסתברות שממוצע אורך המוטות יהיה בין 68 ל 78 ס"מ?

ב. יש לחבר 2 בניינים באמצעות מוטות. המרחק בין שני הבניינים הינו 7200 ס"מ. מה ההסתברות ש 100 המוטות יספיקו למלאכה?

ג. מה צריך להיות גודל המדגם המינימאלי, כדי שבהסתברות של 5% ממוצע המדגם יהיה קטן מ-69 ס"מ. היעזרו במשפט הגבול המרכזי.

9) נתון ש- $X \sim N(\mu, \sigma^2)$. דגמו 5 תצפיות מאותה התפלגות והתבוננו בממוצע

המדגם \bar{X} . לכן: $P(\bar{X} > \mu)$ יהיה (בחרו בתשובה הנכונה):

א. 0.

ב. 0.5.

ג. 1.

ד. לא ניתן לדעת.

10) נתון ש- X מתפלג כלשהו עם תוחלת μ ושונויות σ^2 .

החליטו לבצע מדגם בגודל 200 מתוך ההפלגות הנתונה לפי משפט הגבול המרכזי מתקיים (בחרו בתשובה הנכונה):

א. $X \sim N\left(\mu, \frac{\sigma^2}{200}\right)$

ב. $\mu \sim N\left(\mu, \frac{\sigma^2}{200}\right)$

ג. $\bar{X} \sim N(\mu, \sigma^2)$

ד. $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{200}\right)$

11) נתון ש- $X \sim N(\mu, \sigma^2)$. אם נדגום n תצפיות מתוך ההתפלגות ונגדיר: $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$,

אזי (בחרו בתשובה הנכונה):

א. μ ו- \bar{X} יהיו משתנים מקריים.

ב. μ יהיה משתנה מקרי ו- \bar{X} קבוע.

ג. \bar{X} יהיה משתנה מקרי ו- μ קבוע.

ד. μ ו- \bar{X} יהיו קבועים.

תשובות סופיות:

- (1) א. כלל הסטודנטים במכללה שסיימו סטטיסטיקה א. ב. ציון. ג. ממוצע: 78, סטיית תקן: 15. ד. 2. ה. 78. ו. 10.6.
- (2) א. 0.8413 ב. 0.0013 ג. 0. ד. 0.1974.
- (3) א. 0 ב. 0 ג. 0.9544 ד. 178.205.
- (4) א. 0.0465 ב. 27.71 ג. 0.2628 ד. התשובה הייתה קטנה.
- (5) א. 0 ב. 0.1587 ג. 0.1587 ד. 0.5.
- (6) א. 0.5468 ב. 0.6826
- (7) 0.0475
- (8) א. 0.9772 ב. 0.0228 ג. 271.
- (9) ב'
- (10) ד'
- (11) ג'

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 15 - מושגי יסוד באמידה

תוכן העניינים

1. כללי 50

מושגי יסוד באמידה:

רקע:

כזכור מהמפגש הקודם, פרמטר הוא גודל המתאר את האוכלוסייה או התפלגות מסוימת. כמו ממוצע הגבהים בקרב מתגייסים לצה"ל - μ . כמו פרופורציית התומכים בממשלה בקרב אזרחי המדינה - p . בדרך כלל הפרמטרים הם גדלים שאינם ידועים באמת, ולכן מבצעים מדגמים במטרה לאמוד אותם. אין אפשרות לחשב אותם הניסיון הוא בלהעריך כמה הם שווים ככל שניתן.

- נסמן באופן כללי פרמטר באות θ ואומד ב- $\hat{\theta}$. הוא סטטיסטי המחושב על המדגם ובאמצעותו נאמוד את θ .
- שגיאת אמידה: $|\hat{\theta} - \theta|$ - ההפרש בין האומד לאמת (הפרמטר).

דוגמה (פתרון בהקלטה):

בכנסת ה-19 קיבלה מפלגת העבודה 15 מנדטים. בערוץ 10 ברגע סגירת הקלפיות העריכו את מספר המנדטים של המפלגה להיות 17 מנדטים וזאת על סמך תוצאות מדגם של הערוץ.

- א. מה הפרמטר בדוגמה זו?
 - ב. מהי טעות האמידה של ערוץ 10?
- $E(\hat{\theta}) = \theta$: יהיה אומד חסר הטיה ל- θ אם התוחלת של $\hat{\theta}$ תהיה שווה ל- θ .
 - טעות התקן של אומד היא סטיית התקן שלו, כלומר: $\sigma(\hat{\theta}) = S.E$

פרמטרים מרכזיים והאומדים שלהם:

ממוצע האוכלוסייה μ :

$$\bar{x} = \frac{\sum x}{n} : \text{האומד הנקודתי שלו יהיה: ממוצע המדגם}$$

$$E(\bar{x}) = \mu \text{ , לכן } \bar{x} \text{ הינו אומר חסר הטיה ל-} \mu . \text{ כמו כן, טעות תקן: } \sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = SE$$

פרופורציה באוכלוסייה p :

$$\hat{p} = \frac{y}{n} : \text{האומד הנקודתי שלו יהיה: פרופורציה במדגם}$$

$$E(\hat{p}) = p \text{ , לכן } \hat{p} \text{ הינו אומר חסר הטיה ל-} p . \text{ כמו כן טעות התקן: } \sigma(\hat{p}) = \sqrt{\frac{p \cdot (1-p)}{n}}$$

שונות האוכלוסייה σ^2 :

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} : \text{האומד הנקודתי שלו יהיה}$$

$$E(S^2) = \sigma^2 \text{ ולכן } S^2 \text{ הינו אומד חסר הטיה ל-} \sigma^2 : S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

הערה: אומד הוא הנוסחה הכללית לאמידת הפרמטר ואומדן הוא הערך הספציפי שהתקבל במדגם מסוים.

דוגמה (פתרון בהקלטה):

נדגמו 10 משפחות בתל אביב ונבדק עבור כל משפחה מספר הילדים שלה.
 להלן התוצאות שהתקבלו: 2, 1, 3, 2, 1, 4, 5, 2, 1, 3.
 אמדו באמצעות אומדים חסרי הטיה את הפרמטרים הבאים:

1. ממוצע מספר הילדים למשפחה בתל אביב.
2. שונות מספר הילדים למשפחה בתל אביב.
3. פרופורציית המשפחות בנות שני ילדים.

שאלות:

- (1) מתוך 500 טירונים, נמצאו 120 בעלי שברי הליכה. נתון שהסיכוי שטירון יהיה עם שבר הליכה הוא 0.25.
- מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
 - מהי טעות התקן של האומדן כשהמדגם בגודל 500?
 - מהו האומדן לפרמטר?
 - מהי טעות האמידה?
- (2) לפי נתוני היצרן, מקרר צורך בממוצע 2400 וואט לשעה עם סטיית תקן של 500 וואט לשעה.
- במדגם של 25 מקררים של היצרן התקבל ממוצע של 2342 וואט לשעה.
- מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
 - מהי טעות התקן של האומדן?
 - מהו האומדן לפרמטר?
 - מהי טעות האמידה?
- (3) נדגמו עשרה מתגייסים לצה"ל. גובהם נמדד בס"מ. להלן התוצאות שהתקבלו: 168, 184, 192, 171, 180, 177, 187, 168, 177 ו-175.
- מצאו אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.
 - מצאו אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.
 - מצאו אומדן חסר הטיה לפרופורציות המתגייסים בגובה של לפחות 180 ס"מ.
- (4) נדגמו 20 שכירים באקראי. עבור כל שכיר נמדד השכר באלפי שקלים.
- להלן התוצאות שהתקבלו: $\sum_{i=1}^{20} X_i = 162$, $\sum_{i=1}^{20} X_i^2 = 1502.2$.
- אמדו את השכר הממוצע של השכירים במשק.
 - אמדו את סטיית התקן של שכר השכירים במשק.
- (5) במטרה לאמוד את ממוצע האוכלוסייה, דגמו תצפיות בלתי תלויות מהאוכלוסייה וחישבו את הממוצע שלהם. מהי טעות התקן?
- סטיית התקן של האוכלוסייה.
 - סטיית התקן של ממוצע האוכלוסייה.
 - סטיית התקן של המדגם.
 - סטיית התקן של ממוצע המדגם.

6) משקל הממוצע של אוכלוסייה מסוימת הוא 75 ק"ג עם שונות של 25. אם יבחרו כל המדגמים האפשריים בגודל 10 מאוכלוסייה זו סטיית התקן של ממוצעי המדגמים תהייה:

- א. 3.
- ב. 2.5.
- ג. 1.581.
- ד. אין מספיק נתונים לדעת.

7) במדגם מקרי, מתי סכום ריבועי הסטיות מהממוצע, $\sum_{i=1}^n (x_i - \bar{x})^2$, מחולק ב- $n-1$?

- א. כאשר n קטן.
- ב. כאשר תצפיות המדגם אינן בלתי תלויות.
- ג. כאשר האוכלוסייה אינה מתפלגת נורמאלית.
- ד. כאשר מעוניינים באומדן חסר הטיה לשונות האוכלוסייה ממנה הוצא המדגם.
- ה. כאשר מעוניינים לחשב את שונות התפלגות הדגימה של ממוצע המדגם.

8) X_1, X_2, \dots, X_{16} מדגם מקרי מתוך אוכלוסייה בעלת ממוצע μ לא ידוע ושונות: $\sigma^2 = 64$. טעות התקן של האומדן ל- μ היא:

- א. 16.
- ב. 8.
- ג. 4.
- ד. 2.

9) מהו אומדן חסר הטיה?

- א. אומדן שערכו שווה לממוצע התפלגות הדגימה שלו.
- ב. אומדן שערכו שווה לערך הפרמטר באוכלוסייה.
- ג. אומדן שממוצע התפלגות הדגימה שלו שווה לערך הפרמטר באוכלוסייה.
- ד. אומדן שהסיכוי שערכו יהיה גבוה מערך הפרמטר באוכלוסייה שווה לסיכוי שיהיה נמוך ממנו.

תשובות סופיות:

- (1) א. 0.25 ב. 0.019 ג. 0.24 ד. 0.01
- (2) א. אוכלוסייה: מקררים של יצרן, תוחלת: 2400, סטיית תקן: 500.
 ב. 100 ג. 2342 ד. 58
- (3) א. 177.9 ב. 64.1 ג. 0.4
- (4) א. 8.1 ב. 3.16
- (5) ד'
- (6) ג'
- (7) ד'
- (8) ד'
- (9) ג'

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 16 - מבוא לבדיקת השערות על פרמטרים

תוכן העניינים

55	1. הקדמה
59	2. סוגי טעויות

הקדמה:

רקע:

תהליך של בדיקת השערות הוא תהליך מאד נפוץ בעולם הסטטיסטיקה. בבדיקת השערות על פרמטרים נעבוד לפי השלבים הבאים:

שלב א: נוהה את הפרמטר הנחקר.

שלב ב: נרשום את השערות המחקר.

השערת האפס המסומנות ב- H_0 .

בדרך כלל השערת האפס מסמלת את אשר היה מקובל עד עכשיו, את השגרה, הנורמה.

השערה אלטרנטיבית (השערת המחקר) המסומנת ב- H_1 .

ההשערה האלטרנטיבית מסמלת את החדשנות בעצם ההשערה האלטרנטיבית מדברת על הסיבה שהמחקר נעשה היא שאלת המחקר.

שלב ג: נבדוק האם התנאים לביצוע התהליך מתקיימים ונניח הנחות במידת הצורך.

שלב ד: נרשום את כלל ההכרעה. בתהליך של בדיקת השערות יוצרים כלל שנקרא כלל הכרעה. הכלל יוצר אזורי שנקראים:

1. **אזור דחייה:**

דחייה של השערת האפס כלומר קבלה של האלטרנטיבה.

2. **אזור קבלה:**

קבלה של השערת האפס ודחייה של האלטרנטיבה. כלל ההכרעה מתבסס על איזשהו סטטיסטי. אזור הדחייה מוכתב על ידי סיכון שלוקח החוקר מראש

שנקרא רמת מובהקות ומסומן ב- α .

שלב ה: בתהליך יש ללכת לתוצאות המדגם ולחשב את הסטטיסטי המתאים ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה.

שלב ו: להסיק מסקנה בהתאם לתוצאות המדגם.

דוגמה (פתרון בהקלטה):

משרד הבריאות פרסם שמשקל ממוצע של תינוקות ביום לידתם בישראל 3300 גרם. משרד הבריאות רוצה לחקור את הטענה שנשים מעשנות בזמן ההיריון יולדות תינוקות במשקל נמוך מהממוצע. במחקר השתתפו 20 נשים מעשנות בהריון. להלן תוצאות המדגם שבדק את המשקל של התינוקות בעת הלידה:

$$n = 20, \bar{X} = 3120, S = 280$$

- א. מהי אוכלוסיית המחקר?
- ב. מה המשתנה הנחקר?
- ג. מה הפרמטר הנחקר?
- ד. מהן השערות המחקר?

שאלות:

בשאלות הבאות, ענו על הסעיפים הבאים:

- א. מהי אוכלוסיית המחקר?
- ב. מה המשתנה הנחקר?
- ג. מה הפרמטר הנחקר?
- ד. מהן השערות המחקר?

- (1) ממוצע הציונים בבחינת הבגרות באנגלית הנו 72 עם סטיית תקן 15 נקודות. מורה טוען שפיתח שיטת לימוד חדשה שתעלה את ממוצע הציונים. משרד החינוך החליט לתת למורה 36 תלמידים אקראיים. ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5.
- (2) לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25.
- (3) במשך שנים אחוז המועמדים שהתקבל לפקולטה למשפטים היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. מחקר מעוניין לבדוק האם השנה מקשים על הקבלה לפקולטה למשפטים.
- (4) בחודש ינואר השנה פורסם שאחוז האבטלה במשק הוא 8% במדגם עכשווי התקבל שמתוך 200 אנשים 6.5% מובטלים. רוצים לבדוק ברמת מובהקות של 5% האם כיום אחוז האבטלה הוא כמו בתחילת השנה.

תשובות סופיות:

- (1) א. נבחנים בבגרות באנגלית.
 ב. ציון.
 ג. ממוצע הציונים בשיטת לימוד חדשה.
 ד. $H_0: \mu = 72$
 $H_1: \mu > 72$
- (2) א. משקאות בבקבוק של חברה מסוימת.
 ב. נפח משקה בסמ"ק.
 ג. ממוצע נפח המשקה בבקבוק.
 ד. $H_0: \mu = 500$
 $H_1: \mu < 500$
- (3) א. מועמדים לפקולטה למשפטים.
 ב. משתנה דיכוטומי (התקבל, לא התקבל).
 ג. אחוז הקבלה.
 ד. $H_0: p = 0.25$
 $H_1: p < 0.25$
- (4) א. אזרחים בוגרים במשק.
 ב. משתנה דיכוטומי (מובטל, עובד).
 ג. אחוז האבטלה כיום.
 ד. $H_0: p = 0.08$
 $H_1: p \neq 0.08$

סוגי טעויות:

רקע:

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה. הכלל יוצר אזורים שנקראים:

1. אזור דחייה – דחייה של השערת האפס כלומר קבלה של האלטרנטיבה.
2. אזור קבלה – קבלה של השערת האפס ודחייה של האלטרנטיבה.

כלל ההכרעה מתבסס על איזשהו סטטיסטי. בתהליך יש ללכת לתוצאות המדגם ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה וכך להגיע למסקנה – המסקנה היא בעירבון מוגבל כיוון שהיא תלויה בכלל ההכרעה ובתוצאות המדגם. אם נשנה את כלל ההכרעה אז אנחנו יכולים לקבל מסקנה אחרת. אם נבצע מדגם חדש אז אנחנו עלולים לקבל תוצאה אחרת. לכן יתכנו טעויות במסקנות שלנו:

		הכרעה	
		H_0	H_1
מציאות	H_0	אין טעות	טעות מסוג 1
	H_1	טעות מסוג 2	אין טעות

הגדרת הטעויות:

טעות מסוג ראשון: להכריע לדחות את H_0 למרות שבמציאות H_0 נכונה.

טעות מסוג שני: להכריע לקבל את H_0 למרות שבמציאות H_1 נכונה.

דוגמה (פתרון בהקלטה):

אדם חשוד בביצוע עבירה ונתבע בבית המשפט. אילו סוגי טעויות אפשריות בהכרעת הדין?

שאלות:

- (1) לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25. בסופו של דבר הוחלט להכריע לטובת חברת המשקאות.
- א. רשמו את השערות המחקר.
 ב. מה מסקנת המחקר?
 ג. איזו סוג טעות יתכן וביצעו במחקר?
- (2) במחקר על פרמטר מסוים הוחלט בסופו של דבר לדחות את השערת האפס.
- א. האם ניתן לדעת אם בוצע טעות במחקר?
 ב. מה סוג הטעות האפשרית?
- (3) לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. ישנה טענה שכיום ממוצע מספר הילדים במשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה. על סמך תוצאות המדגם נקבע שלא ניתן לקבוע שבאופן מובהק תוחלת מספר הילדים למשפחה קטנה כיום.
- א. מהי אוכלוסיית המחקר?
 ב. מה המשתנה הנחקר?
 ג. מה הפרמטר הנחקר?
 ד. מה השערות המחקר?
 ה. מה מסקנת המחקר?
 ו. מהי סוג הטעות האפשרית במחקר?

תשובות סופיות:

- (1) א. $H_0: \mu = 500$
 ב. לא דחינו את H_0 .
 ג. טעות מסוג שני.
- (2) א. לא ניתן לדעת.
 ב. טעות מסוג ראשון.
 (3) א. משפחות כיום.
 ב. מס' הילדים.
 ג. תוחלת מספר הילדים למשפחה כיום.
 ה. לא לדחות את H_0 . ו. טעות מסוג שני.
- ד. $H_0: \mu = 2.3$
 $H_1: \mu < 2.3$

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

פרק 17 - רווח סמך לתוחלת (ממוצע)

תוכן העניינים

1. רווח סמך כששונות האוכלוסיה ידועה 61
2. רווח סמך כששונות האוכלוסיה לא ידועה 66

רווח סמך כששונות האוכלוסייה ידועה:

רקע:

ממוצע המדגם הוא אומדן לממוצע האוכלוסייה, אך לא באמת ניתן להבין ממנו על גודלו של ממוצע האוכלוסייה. ההסתברות שממוצע המדגם יהיה בדיוק כמו הממוצע האמתי הוא אפסי.

מה שנהוג לעשות כדי לאמוד את ממוצע האוכלוסייה, זה לבנות רווח סמך.

בנה מרווח בטחון שהסיכוי שהפרמטר μ ייכלל בתוכו הוא: $1-\alpha$.

$1-\alpha$: נקרא רמת בטחון או רמת סמך. כך ש: $P(A \leq \mu \leq B) = 1-\alpha$.

A - גבול התחתון של רווח הסמך.

B - הגבול העליון של רווח הסמך.

$L = B - A$ - אורך רווח הסמך.

דוגמה (פתרון בהקלטה):

חוקר דגם 25 חיילים שנבחנו במבחן הפסיכומטרי. הוא בנה רווח סמך לממוצע הציונים במבחן הפסיכומטרי בקרב אוכלוסיית החיילים וקיבל בין 510 ל-590. רווח הסמך נבנה ברמת סמך של 95%.

1. מהי אוכלוסיית המחקר?
2. מה המשתנה באוכלוסייה?
3. מה הפרמטר שהחוקר רצה לאמוד?
4. מהו רווח הסמך?
5. מה אורך רווח הסמך?
6. מהי רמת הביטחון של רווח הסמך?

בפרק זה נרצה לבנות רווח סמך לתוחלת (μ) במקרה ש- σ^2 (שונות האוכלוסייה) ידועה.

פרמטר אותו נרצה לאמוד: μ .

אומד נקודתי: \bar{x} .

תנאים לבניית רווח הסמך: $X \sim N$ או $n \geq 30$.

σ^2 (שונות האוכלוסייה) ידועה.

נוסחה לרווח הסמך: $\bar{x} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

דוגמה (פתרון בהקלטה):

על פי נתוני היצרן אורך חיי סוללה מתפלג נורמאלית עם סטיית תקן של 1 שעה. מעוניינים לאמוד את תוחלת חיי סוללה. נדגמו באקראי 4 סוללות, אורך החיים הממוצע שהתקבל הוא 13.5 שעות. בנו רווח סמך ברמת סמך של 95% לתוחלת אורך חיי סוללה.

שגיאת האמידה המקסימלית: $\varepsilon = Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

ε - נותן את שגיאת האמידה המקסימלית, דבר שנקרא גם טעות סטטיסטית, טעות דגימה.

דוגמה (פתרון בהקלטה):

בהמשך לשאלה עם הסוללות. מה ניתן להגיד בביטחון של 95% על שגיאת האמידה?

קשרים מתמטיים ברווח הסמך:

• אורך רווח הסמך הוא פעמיים שגיאת האמידה המקסימלית: $L = 2\varepsilon$.

• ממוצע המדגם נופל תמיד באמצע רווח הסמך: $\bar{X} = \frac{A+B}{2}$.

• ככל שמספר התצפיות (n) גבוה יותר, כך יש יותר אינפורמציה ולכן האומד יותר מדויק, ולכן נקבל רווח סמך יותר קצר.

• ככל שרמת הביטחון ($1-\alpha$) גבוהה יותר, כך: $z_{1-\frac{\alpha}{2}}$ יותר גבוה, ורווח הסמך יותר ארוך.

שאלות:

- 1) חוקר התעניין לאמוד את השכר הממוצע במשק. על סמך מדגם הוא קבע שבביטחון של 95% כי השכר הממוצע במשק נע בין 9200 ל-9800 ₪.
- מי האוכלוסייה במחקר?
 - מה המשתנה הנחקר?
 - מה הפרמטר שאותו רוצים לאמוד?
 - מה רווח הסמך לפרמטר?
 - מהי רמת הסמך לפרמטר?
 - מה אורך רווח הסמך?
 - מה הסיכוי שטעות הדגימה תעלה על 300 ₪?
- 2) מעוניינים לאמוד את התפוקה היומית הממוצעת של מפעל מסוים ברמת סמך של 95%. במדגם אקראי של 100 ימים התקבלה תפוקה ממוצעת 4950 מוצרים ביום. לצורך פתרון הנח שסטיית התקן האמתית ידועה ושווה 150 מוצרים ביום. בנו את רווח הסמך.
- 3) מעוניינים לאמוד את ממוצע אורך החיים של מכשיר. מנתוני היצרן ידוע שאורך החיים מתפלג נורמאלי עם סטיית תקן של 20 שעות. נדגמו 25 מכשירים ונמצא כי ממוצע אורך החיים שלהם היה 230 שעות.
- בנו רווח סמך ברמת סמך של 90% לאורך החיים הממוצע של מכשיר.
 - בנו רווח סמך ברמת סמך של 95% לאורך החיים הממוצע של מכשיר.
 - הסבירו כיצד ומדוע השתנה רווח הסמך.
- 4) דגמו 200 עובדים מהמשק הישראלי. השכר הממוצע שלהם היה 9700 ₪. נניח שסטיית התקן של השכר במשק היא 3000 ₪.
- בנו רווח סמך ברמת סמך של 95% לתוחלת השכר במשק.
 - מה ניתן לומר בביטחון של 95% על הסטייה המרבית בין ממוצע המדגם לתוחלת השכר?
 - מה היה צריך להיות גודל המדגם אם הינו רוצים להקטין את רווח הסמך ב-50%?
 - אם היינו מגדילים את גודל המדגם ובונים רווח סמך באותה רמת סמך האם היה ניתן לטעון בביטחון רב יותר שרווח הסמך מכיל את הפרמטר?

- (5) בנו רווח סמך לממוצע הציונים של מבחן אינטליגנציה. ידוע שסטיית התקן היא 15 והמדגם מתבסס על 100 תצפיות. רווח הסמך שהתקבל הוא (105,99). שחזרו את:
- ממוצע המדגם.
 - שגיאת האמידה המקסימאלית.
 - רמת הסמך.
- (6) זמן החלמה מאנגינה מתפלג עם סטיית תקן של יומיים. חברת תרופות מעוניינת לחקור אנטיביוטיקה חדשה שהיא פיתחה. במחקר השתתפו 60 אנשים שחלו באנגינה וקיבלו את האנטיביוטיקה החדשה. בממוצע הם החלימו לאחר 4 ימים.
- בנו רווח סמך לתוחלת זמן ההחלמה תחת האנטיביוטיקה החדשה ברמת סמך של 90%.
 - מה היה קורה לאורך רווח הסמך אם היה תקציב להגדלת גודל המדגם פי 4? הסבירו.
 - מה היה קורה לאורך רווח הסמך אם היינו בונים את רווח הסמך ברמת סמך גדולה יותר? הסבירו.
- (7) חוקר בנה רווח סמך לממוצע וקיבל את רווח הסמך הבא: $82 < \mu < 92$. נתון שסטיית התקן בהתפלגות שווה ל-10 ושהמדגם מתבסס על 16 תצפיות. התפלגות המשתנה היא נורמאלית.
- מהו ממוצע המדגם?
 - מהי רמת הסמך של רווח הסמך שנבנה?
 - מה הסיכוי ששגיאת האמידה באמידת ממוצע האוכלוסייה תעלה על 5?
- (8) חוקר בנה רווח סמך לתוחלת כאשר השונות בהתפלגות ידועה ברמת סמך של 95%. אם החוקר כעת יבנה על סמך אותם נתונים רווח סמך ברמת סמך קטנה מ-95%, איזה מהמשפטים הבאים לא יהיה נכון.
- אורך רווח הסמך החדש יהיה קטן יותר.
 - גודל המדגם יהיה כעת קטן יותר.
 - המרחק בין ממוצע המדגם לקצות רווח הסמך יהיו קטנים יותר ברווח הסמך החדש.
 - רמת הביטחון לבנות רווח הסמך החדש תהיה קטנה יותר.

(9) חוקר בנה רווח סמך ל- μ וקיבל: $48 < \mu < 54$. מה נכון בהכרח:

א. $\mu = 51$.

ב. $\bar{X} = 6$.

ג. $\bar{X} = 51$.

ד. אורך רווח הסמך הינו 3.

(10) איזה מהגורמים הבאים אינו משפיע על גודלו של רווח בר סמך, כאשר שונות האוכלוסייה ידועה (בחרו בתשובה הנכונה):

א. רמת הביטחון.

ב. סטיית התקן באוכלוסייה.

ג. מספר המשתתפים.

ד. סטיית התקן במדגם.

תשובות סופיות:

(1) א. העובדים במשק. ב. שכר ב-ש. ג. μ . ד. $9200 < \mu < 9800$.

ז. 0.05.

ו. 600.

ה. 0.95.

(2) $4920.6 < \mu < 4979.4$

ב. $222.16 < \mu < 237.84$

(3) א. $223.42 < \mu < 236.58$

ג. ראה סרטון.

(4) א. $10,116 < \mu < 9284$. ב. הסטייה המירבית בין \bar{x} ל- μ היא 416 שם בבטחון של 95%.

ד. לא.

ג. 800.

ג. 0.9544.

ב. 3.

(5) א. 102.

ג. גדל.

ב. יקטן פי 2.

(6) א. $4.42 < \mu < 83.5$.

ג. 0.9544.

ב. 5.

(7) א. 87.

(8) ב'.

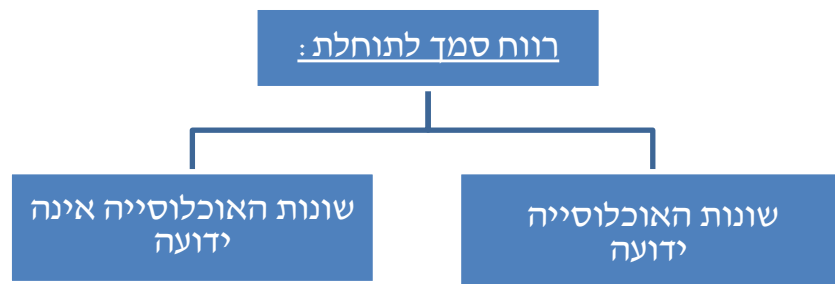
(9) ג'.

(10) ד'.

רווח סמך כששונות האוכלוסייה לא ידועה:

רקע:

בבואנו לבנות רווח סמך לתוחלת אנו צריכים להתמקד בשני המצבים הבאים:

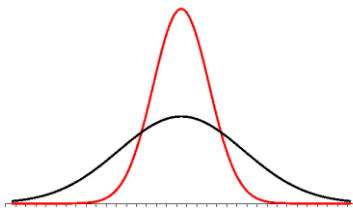


בפרק זה נעסוק במקרה ששונות האוכלוסייה (σ^2) אינה ידועה לנו.

מקרה יותר פרקטי.

התנאי: $X \sim N$ או שהמדגם גדול.

רווח סמך: $\bar{X} \pm t_{1-\frac{\alpha}{2}}^{(n-1)} \cdot \frac{S}{\sqrt{n}}$



$$\text{האומד לשונות: } S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$$

התפלגות T:

הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה

להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים.

התפלגות T תלויה במושג שנקרא דרגות חופש. דרגות החופש הן: $df = n-1$.

ככל שדרגות החופש עולות ההתפלגות הופכת להיות יותר גבוהה וצרה.

כשדרגות החופש שואפות לאינסוף התפלגות T שואפת להיות כמו התפלגות Z.

דוגמה (פתרון בהקלטה):

הזמן שלוקח לפתור שאלה מסוימת בחשבון מתפלג אצל תלמידי כיתות ח' נורמאלית.

במטרה לאמוד את תוחלת זמן הפתרון נדגמו 4 תלמידים בכיתה ח'. להלן התוצאות

שהתקבלו בדקות: 4.7, 5.2, 4.6, 5.3.

בנו רווח סמך ברמת סמך של 95% לממוצע זמן הפתרון לשאלה בקרב תלמידי כיתה ח'.

שאלות:

- (1) מחקר מעוניין לדעת כיצד תרופה מסוימת משפיעה על קצב פעימות הלב. ל-5 אנשים שנטלו את התרופה מדדו את הדופק והתקבל מספר פעימות לדקה: 84, 88, 84, 79, 89. הערה: לצורך פתרון הנח שקצב פעימות הלב מתפלג נורמאלית בקירוב.
- א. בנו רווח סמך ברמת סמך של 95% לתוחלת הדופק של נוטלי התרופה הנ"ל.
 ב. נתון שהדופק הממוצע ללא לקיחת התרופה הינו 70. לאור זאת, האם בביטחון של 95% התרופה משפיעה על הדופק?
 ג. בהמשך לסעיף א', אם היינו בונים את רווח הסמך ברמת ביטחון של 99%, כיצד הדבר היה משפיע על רווח הסמך?
- (2) במדגם שנעשה על 25 מתגייסים לצבא האמריקאי התקבל כי גובה ממוצע של חייל הינו 178 ס"מ עם סטיית תקן: $S = 13$ ס"מ. בנו רווח סמך ברמת סמך של 90% לתוחלת גובה המתגייסים לצבא האמריקאי. מה יש להניח לצורך פתרון?
- (3) אדם מעוניין לאמוד את זמן הנסיעה הממוצע שלו לעבודה. לצורך כך הוא דוגם 5 ימים שזמן הנסיעה בהם בדקות הוא: 30, 40, 32, 34, 27. א. ברמת ביטחון של 95% אמוד את זמן הנסיעה הממוצע. מהי ההנחה הדרושה לצורך פתרון?
 ב. איך גודל רווח הסמך היה משתנה אם היו דוגמים עוד ימים?
- (4) ציוני מבחן אינטליגנציה מתפלגים נורמאלית. נדגמו 25 מבחנים והתקבל ממוצע ציונים 102 וסטיית תקן מדגמית 13. א. בנו רווח סמך לממוצע הציונים באוכלוסייה ברמת ביטחון של 95%.
 ב. חזרו על סעיף א' אם סטיית התקן הינה סטיית התקן האמתית של כלל הנבחנים.
 ג. הסבירו את ההבדלים בין שני הסעיפים הנ"ל.
- (5) נשקלו 60 תינוקות אשר נולדו בשבוע ה-40 של ההיריון. המשקל נמדד בקילוגרמים. להלן התוצאות שהתקבלו: $\sum_{i=1}^{60} X_i = 195$, $\sum_{i=1}^{60} X_i^2 = 643.19$. בנו רווח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היוולדו.

- (6) נדגמו 120 אנשים אקראיים מעל גיל 50. עבור כל אדם נבדק מספר שנות השכלתו. להלן התוצאות שהתקבלו: $\bar{x} = 13.8$, $S = 2$. בנו רווח סמך ברמת סמך של 96% לממוצע ההשכלה של אזרחים מעל גיל 50.
- (7) שני סטטיסטיקאים בנו רווח בר-סמך לאותו פרמטר μ . לכל אחד מהסטטיסטיקאים מדגם אחר, אך באותו גודל 10. שניהם קבעו אותה רמת סמך. סטטיסטיקאי א': הניח $\sigma = 20$. סטטיסטיקאי ב': חישב לפי המדגם וקיבל $S = 20$. למי משני הסטטיסטיקאים יהיה רווח סמך ארוך יותר?
 א. סטטיסטיקאי א'.
 ב. סטטיסטיקאי ב'.
 ג. אותו אורך רווח סמך לשני הסטטיסטיקאים.
 ד. תלוי בתוצאות המדגם של כל סטטיסטיקאי.
- (8) נתון ש: $X \sim N(\mu, \sigma^2)$ ביצעו מדגם בגודל 16 וקיבלו סטיית תקן מדגמית 10. אורך רווח הסמך שהתקבל הוא: 8.765. מהי רמת הביטחון של רווח הסמך?

תשובות סופיות:

- (1) א. $79.88 < \mu < 89.72$ ב. כן. ג. הוא היה גדל.
- (2) ראה בסרטון.
- (3) א. צריך להניח שהמשתנה מתפלג נורמלית. ב. לא ניתן לדעת.
- (4) א. $96.63 < \mu < 107.37$ ב. $96.90 < \mu < 107.10$ ג. ראה בסרטון.
- (5) $3.149 < \mu < 3.351$
- (6) $13.42 < \mu < 14.18$
- (7) ב'.
- (8) 90%

מבוא לסטטיסטיקה לפסיכולוגים - סמסטר א

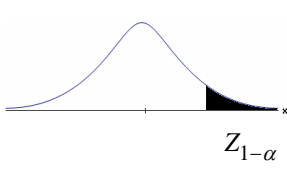
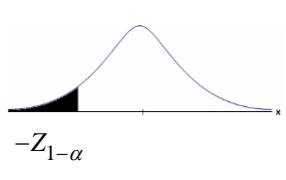
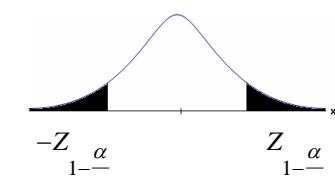
פרק 18 - בדיקת השערות על תוחלת (ממוצע)

תוכן העניינים

1. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה ידועה. 69
2. מובהקות תוצאה - אלפא מינימלית (ששונות האוכלוסיה ידועה) 73
3. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה לא ידועה. 78
4. ניתוח פלטים. 82
5. הקשר בין רווח סמך לבדיקת השערות על תוחלת (ממוצע). 87

בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסייה ידועה:

רקע:

$H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	השערת האפס: השערה אלטרנטיבה:
1. σ ידועה 2. $X \sim N$ או מדגם מספיק גדול			תנאים:
$Z_{\bar{x}} > Z_{1-\alpha}$	$Z_{\bar{x}} < -Z_{1-\alpha}$	$Z_{\bar{x}} < -Z_{1-\frac{\alpha}{2}}$ או $Z_{\bar{x}} > Z_{1-\frac{\alpha}{2}}$	כלל ההכרעה: אזור הדחייה של H_0 :
			
דוחים את H_0 ■	דוחים את H_0 ■	דוחים את H_0 ■	

$$Z_{\bar{x}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

סטטיסטי המבחן:

חלופה אחרת לכלל הכרעה:

$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ או $\bar{X} < \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$	נדחה H_0 אם מתקיים:
--	--	--	--------------------------

דוגמה:

יבול העגבניות מתפלג נורמלית עם תוחלת של 10 טון לדונם וסטיית תקן של 2.5 טון לדונם בעונה. משערים ששיטת זיבול חדשה תעלה את תוחלת היבול לעונה מבלי לשנות את סטיית התקן. נדגמו 4 חלקות שזובלו בשיטה החדשה. היבול הממוצע שהתקבל היה 12.5 טון לדונם. בדקו את ההשערה ברמת מובהקות של 1%.

פיתרון:

אוכלוסייה: עגבניות.

המשתנה: $X =$ יבול העגבניות בטון לעונה.

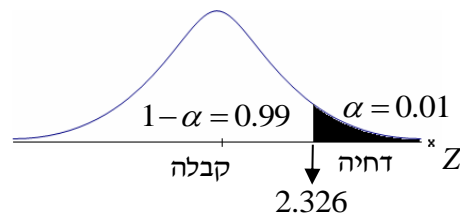
הפרמטר: $\mu =$ תוחלת היבול בשיטת הזיבול החדשה.

השערות:
 $H_0: \mu = 10$
 $H_1: \mu > 10$

תנאים:

1. $X \sim N$

2. $\sigma = 2.5$

כלל הכרעה:

נדחה את H_0 אם $Z_{\bar{x}} > 2.326$

תוצאות: $n = 4$, $\bar{x} = 12.5$

סטטיסטי המבחן: $Z_{\bar{x}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

נציב: $Z_{\bar{x}} = \frac{1.25 - 10}{\frac{2.5}{\sqrt{4}}} = 2 < 2.326$

מסקנה:

לא נדחה H_0 (נקבל H_0).

ברמת מובהקות של 1% לא נוכל לקבל את הטענה ששיטת הזיבול החדשה מעלה את תוחלת היבול של העגבניות.

שאלות:

- (1) ממוצע הציונים בבחינת הבגרות באנגלית הנו 72 עם סטיית תקן 15 נקודות. מורה טוען שפיתח שיטת לימוד חדשה שתעלה את ממוצע הציונים. משרד החינוך החליט לתת למורה 36 תלמידים אקראיים. ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5. בהנחה שגם בשיטתו סטיית התקן תהיה 15 מה מסקנתכם ברמת מובהקות של 5%?
- (2) לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 ס"מ³ וסטיית תקן 20 ס"מ³. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 ס"מ³ במדגם בגודל 25. א. מה מסקנתכם ברמת מובהקות של 2.5%? ב. האם ניתן לדעת מה תהיה המסקנה עבור רמת מובהקות הגבוהה מ-5%?
- (3) מהנדס האיכות מעוניין לבדוק אם מכונה מכיילת (מאופסת). המכונה כוונה לחתוך מוטות באורך 50 ס"מ. לפי נתוני היצרן סטיית התקן בחיתוך המוטות היא 0.5 ס"מ. במדגם של 50 מוטות התקבל ממוצע אורך המוט 50.93 ס"מ. מה מסקנתכם ברמת מובהקות של 5%?
- (4) המשקל הממוצע של הספורטאים בתחום ספורט מסוים הוא 90 ק"ג, עם סטיית תקן 8 ק"ג. לפי דעת מומחים בתחום יש צורך בהורדת המשקל ובשימוש בדיאטה מסוימת שצריכה להביא להורדת המשקל. לשם בדיקת יעילות הדיאטה נלקח מדגם מקרי של 50 ספורטאים ובתום שנה של שימוש בדיאטה התברר שהמשקל הממוצע במדגם זה היה 84 ק"ג. יש לבדוק בר"מ של 10%, האם הדיאטה גורמת להורדת המשקל.
- (5) לפי מפרט נתון, על עובי בורג להיות 4 מ"מ עם סטיית תקן של 0.2 מ"מ. במדגם של 25 ברגים העובי הממוצע היה 4.07 מ"מ. קבעו ברמת מובהקות 0.05, האם עובי הברגים מתאים למפרט. הניחו כי עובי של בורג מתפלג נורמלית וסטיית התקן של עובי בורג היא אכן 0.2 מ"מ.
- (6) במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5% מה תמיד נכון? בחרו בתשובה הנכונה.
- א. הגדלת רמת המובהקות לא תשנה את מסקנת המחקר.
 ב. הגדלת רמת המובהקות תשנה את מסקנת המחקר.
 ג. הקטנת רמת המובהקות לא תשנה את מסקנת המחקר.
 ד. הקטנת רמת המובהקות תשנה את מסקנת המחקר.

(7) חוקר ערך מבחן דו צדדי ברמת מובהקות של α והחליט לדחות את השערת האפס.

אם החוקר היה עורך מבחן צדדי ברמת מובהקות של $\frac{\alpha}{2}$ אזי בהכרח:

- א. השערת האפס הייתה נדחית.
- ב. השערת האפס הייתה לא נדחית.
- ג. לא ניתן לדעת מה תהיה מסקנתו במקרה זה.

(8) שני סטטיסטיקאים בדקו השערות: $H_0: \mu = \mu_0$ כנגד $H_1: \mu > \mu_0$,

עבור שונות ידועה ובאותה רמת מובהקות.

שני החוקרים קבלו אותו ממוצע במדגם אך לחוקר א' היה מדגם בגודל 100 ולחוקר ב' מדגם בגודל 200.

- א. אם חוקר א' החליט לדחות את H_0 , מה יחליט חוקר ב'? נמקו.
- ב. אם חוקר א' יחליט לא לדחות את H_0 , מה יחליט חוקר ב'? נמקו.

תשובות סופיות:

(1) נקבל H_0 , בר"מ של 5% לא נקבל את הטענה של המורה ששיטת הלימוד שלו מעלה את ממוצע הציונים.

(2) א. נדחה H_0 , בר"מ של 2.5% נקבל את תלונת אגודת הצרכנים בדבר הפחתת נפח המשקה בבקבוק.

ב. הגדלנו את רמת המובהקות לכן אנחנו נשארים בדחייה של H_0 והמסקנה לא תשתנה.

(3) נדחה H_0 , בר"מ של 5% נקבע שהמכונה לא מאופסת.

(4) נדחה H_0 , בר"מ של 0.1 נקבל את הטענה שהדיאטה יעילה ומפחיתה את המשקל הממוצע.

(5) נקבל H_0 , בר"מ של 0.05 נכריע שתוחלת עובי הבורג מתים למפרט.

(6) א'.

(7) ג'.

(8) א. לדחות. ב. לא ניתן לדעת.

מובהקות תוצאה – אלפא מינימלית (ששונות האוכלוסייה ידועה):

רקע:

דרך נוספת להגיע להכרעות שלא דרך כלל הכרעה, היא דרך חישוב מובהקות התוצאה:

באמצעות תוצאות המדגם מחשבים את מובהקות התוצאה שמסומן ב- p_v . את רמת המובהקות החוקר קובע מראש לעומת זאת, את מובהקות התוצאה החוקר יוכל לחשב רק אחרי שיהיו לו את התוצאות.

המסקנה של המחקר תקבע לפי העיקרון הבא: אם $p_v \leq \alpha$, דוחים את H_0 . מובהקות התוצאה זה הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

(לקבל את תוצאות המדגם וקיצוני) $p_v = P_{H_0}$.

אם ההשערה היא דו צדדית:

(לקבל את תוצאות המדגם וקיצוני) $p_v = 2P_{H_0}$.

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	השערת האפס: השערה אלטרנטיבה:
1. σ ידועה			תנאים:
2. $X \sim N$ או מדגם מספיק גדול			
$P_{H_0}(\bar{X} \geq \bar{x})$	$P_{H_0}(\bar{X} \leq \bar{x})$	אם $2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \leftarrow \bar{x} > \mu_0$ אם $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \leftarrow \bar{x} < \mu_0$	p-value

כאשר בהנחת השערת האפס: $Z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$, $\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$

דוגמה:

המשקל הממוצע של מתגייסים לצבא לפני 20 שנה היה 65 ק"ג. מחקר מעוניין לבדוק האם כיום המשקל הממוצע של מתגייסים גבוה יותר. נניח שמשקל המתגייסים מתפלג נורמאלית עם סטיית תקן של 12 ק"ג. במדגם של 16 מתגייסים התקבל משקל ממוצע של 71 ק"ג.

א. מהי מובהקות התוצאה?

ב. מה המסקנה אם רמת המובהקות היא 5% ואם רמת המובהקות היא 1%?

פתרון:

א. אוכלוסייה: המתגייסים לצבא כיום.

משתנה: $X =$ משקל בק"ג.

פרמטר: μ .

$$H_0: \mu = 65$$

השערות: $H_1: \mu > 65$

תנאים:

$$1. X \sim N$$

$$2. \sigma = 12$$

תוצאות מדגם:

$$n = 16$$

$$\bar{X} = 71$$

$$P_V = P_{H_0} \left(\begin{array}{c} \text{לתוצאות} \\ \text{המדגם} \\ \text{וקיצוני} \end{array} \right) = P_{H_0} (\bar{X} \geq 71) = 1 - \phi(2) = 1 - 0.9772 = 0.0228$$

$$Z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{71 - 65}{\frac{12}{\sqrt{16}}} = 2$$

$$\alpha_{\min} = 0.0228$$

שאלות:

- (1) להלן השערות של מחקר: $H_0: \mu = 70$, $H_1: \mu > 70$. המשתנה הנחקר מתפלג נורמלית עם סטיית תקן 20. במדגם מאותה אוכלוסייה התקבלו התוצאות הבאות: $n = 100$, $\bar{x} = 74$. מהי מובהקות התוצאה?
- (2) השכר הממוצע במשק בשנת 2012 היה 8800 ₪ עם סטיית תקן 2000. במדגם שנעשה אתמול על 100 עובדים התקבל שכר ממוצע 9500 ₪. מטרת המחקר היא לבדוק האם כיום חלה עליה בשכר. עבור אילו רמות מובהקות שיבחר החוקר יוחלט שחלה עליה בשכר הממוצע במשק?
- (3) אדם חושד שחברת ממתקים לא עומדת בהתחייבויותיה, ומשקלו של חטיף מסוים אותו הוא קונה מדי בוקר נמוך מ-100 גרם. חברת הממתקים טוענת מצידה שהיא אכן עומדת בהתחייבויותיה. ידוע כי סטיית התקן של משקל החטיף היא 12 גרם. האדם מתכוון לשקול 100 חפיסות חטיפים ולאחר מכן להגיע להחלטה. לאחר הבדיקה הוא קיבל משקל הממוצע של 98.5 גרם.
- א. רשמו את השערות המחקר.
 ב. מהי רמת המובהקות המינימלית עבורה דוחים את השערת האפס?
 ג. מהי רמת המובהקות המקסימלית עבורה נקבל את השערת האפס?
 ד. מה המסקנה ברמת מובהקות של 5?
- (4) מכונה לחיתוך מוטות במפעל חותכת מוטות באורך שמתפלג נורמלית עם תוחלת אליה כוונה המכונה וסטיית תקן 2 ס"מ. ביום מסוים כוונה המכונה לחתוך מוטות באורך 80 ס"מ. אחראי האיכות מעוניין לבדוק האם המכונה מכוילת. לצורך כך נדגמו מקו הייצור 16 מוטות שנחתכו אורכן הממוצע היה 81.7 ס"מ.
- א. מהי רמת המובהקות המינימלית עבורה נכריע שהמכונה לא מכוילת?
 ב. אם נוסיף עוד תצפית שערכה יהיה 82 ס"מ, כיצד הדבר ישפיע על התשובה של הסעיף הקודם?
 ג. הכרע ברמת מובהקות של 5% האם המכונה מכוילת.
- (5) אם מקבלים בחישובים אלפא מינימלית (P value) קטנה מאד, סביר להניח כי החוקר ידחה את השערת האפס בקלות. נכון/לא נכון? נמק.

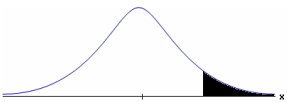

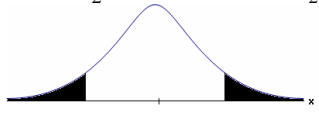
- 6) בבדיקת השערות התקבל שה- $p\text{-value} = 0.02$.
 מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%?
 בחרו בתשובה הנכונה.
- א. יקבל את השערת האפס בכל מקרה.
 - ב. ידחה את השערת האפס מקרה.
 - ג. ידחה את השערת האפס רק אם המבחן הנו דו צדדי.
 - ד. לא ניתן לדעת כי אין מספיק נתונים.
- 7) מובהקות התוצאה (PV) היא גם (בחרו בתשובה הנכונה):
- א. רמת המובהקות המינימאלית לדחות השערת האפס.
 - ב. רמת המובהקות המקסימאלית לדחיית השערת האפס.
 - ג. רמת המובהקות שנקבעת מראש על ידי החוקר שטרם קיבל את תוצאות המחקר.
 - ד. רמת המובהקות המינימאלית לאי דחיית השערת האפס.
- 8) בבדיקת השערות מסוימת התקבל: $p\text{ value} = 0.0254$ לכן (בחרו בתשובה הנכונה):
- א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את H_0 .
 - ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את H_0 .
 - ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את H_0 .
 - ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את H_0 .

תשובות סופיות:

- (1) 0.0228.
- (2) עבור כל רמת מובהקות סבירה.
- (3) א. $H_0: \mu = 100$
 $H_1: \mu < 100$
 ב. 0.1056. ג. 0.1056.
- ד. נכריע שיש עמידה בהתחייבות של החברה.
- (4) א. 0.0006. ב. יקטן. ג. נכריע שאין כיול.
- (5) נכון.
- (6) א'.
- (7) א'.
- (8) ג'.

בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסייה לא ידועה:

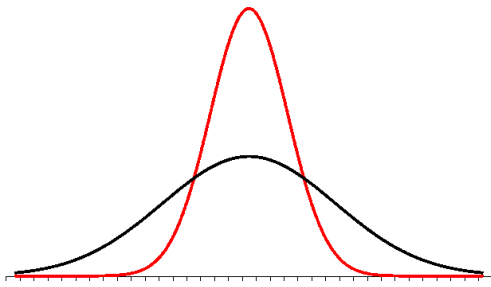
רקע:

$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	השערת האפס: השערה אלטרנטיבה:
1. σ אינה ידועה 2. $X \sim N$ או מדגם מספיק גדול			תנאים:
$t_{\bar{x}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$ - דוחים את H_0	$t_{\bar{x}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha, n-1}$ - דוחים את H_0	$t_{\bar{x}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{x}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$  $-t_{1-\frac{\alpha}{2}, n-1}$ $t_{1-\frac{\alpha}{2}, n-1}$ - דוחים את H_0	כלל ההכרעה: אזור הדחייה של H_0 :
$\bar{X} > \mu_0 + t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} < \mu_0 - t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} > \mu_0 + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$ או $\bar{X} < \mu_0 - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$	חלופה לכלל הכרעה: נדחה H_0 אם מתקיים:

$$t_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}}$$

סטטיסטי המבחן:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$$



התפלגות T:

הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים. התפלגות T תלויה במושג שנקרא דרגות חופש.

דרגות החופש הן: $df = n - 1$.

ככל שדרגות החופש עולות ההתפלגות הופכת להיות יותר גבוהה וצרה. כשדרגות החופש שואפות לאינסוף התפלגות T שואפת להיות כמו התפלגות Z.

דוגמה (פתרון בהקלטה):

מפעל קיבל הזמנה לייצור משטחים בעובי של 0.1 ס"מ. כדי לבדוק האם המפעל עומד בדרישה נדגמו 10 משטחים ונמצא שהעובי הממוצע הוא 0.104 עם אומדן לסטיית תקן 0.002 ס"מ.

א. מהן השערות המחקר?

ב. מה ההנחה הדרושה לצורך פתרון?

ג. בדוק ברמת מובהקות של 5%.

שאלות:

(1) משך זמן ההחלמה בלקיחת אנטיביוטיקה מסוימת הוא 120 שעות בממוצע עם סטיית תקן לא ידועה. מעוניינים לבדוק האם אנטיביוטיקה אחרת מקטינה את משך זמן ההחלמה. במדגם של 5 חולים שלקחו את האנטיביוטיקה האחרת התקבלו זמני ההחלמה הבאים: 90, 95, 100, 80, 125 שעות. מה מסקנתכם ברמת מובהקות של 5%. מהי ההנחה הדרושה לצורך הפתרון?

(2) משרד הבריאות פרסם שמשקל ממוצע של תינוקות ביום היוולדם בישראל 3300 גר'. משרד הבריאות רוצה לחקור את הטענה ששנים מעשנות בזמן ההיריון יולדות תינוקות במשקל נמוך מהממוצע. במחקר השתתפו 20 נשים מעשנות בהריון. להלן תוצאות המדגם שבדק את המשקל של התינוקות בעת הלידה:

$$n = 20$$

$$\bar{x} = 3120$$

$$S = 280$$

מה מסקנתכם ברמת מובהקות של 5% מה יש להניח לצורך פתרון?

(3) ציוני מבחן אינטליגנציה מתפלגים נורמלית. בארה"ב ממוצע הציונים הוא 100. במדגם שנעשה על 23 נבחנים ישראלים, התקבל ממוצע ציונים 104.5 וסטיית התקן המדגמית 16. האם בישראל ממוצע הציונים שונה מארה"ב? הסיקו ברמת מובהקות של 5%.

(4) באוכלוסייה מסוימת נדגמו 10 תצפיות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{10} X_i = 750$$

$$\sum_{i=1}^{10} (X_i - \bar{X})^2 = 900$$

נתון שההתפלגות היא נורמלית.

בדוק ברמת מובהקות של 5% האם התוחלת של ההתפלגות שונה מ-80.

- (5) ליאור ורוני העלו את אותן השערות על ממוצע האוכלוסייה. כמו כן הם התבססו על אותן תוצאות של מדגם. ליאור השתמש בטבלה של התפלגות Z . רוני השתמשה בטבלה של התפלגות t . מה נוכל לומר בנוגע להחלטת המחקר שלהם? בחר בתשובה הנכונה.
- אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.
 - אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.
 - שני החוקרים בהכרח יגיעו לאותה מסקנה.
 - לא ניתן לדעת על היחס בין דחיית השערת האפס של שני החוקרים.

- (6) נתון ש: $X \sim N(\mu, \sigma^2)$ כמו כן נתונות ההשערות הבאות: $H_0: \mu = \mu_0$, $H_1: \mu < \mu_0$

חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות. σ^2 לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לחזק את קביעתו הוא דגם עוד 5 תצפיות ושקלל את תוצאות אלה גם למדגם כך שכלל עכשיו 15 תצפיות. בחר בתשובה הנכונה:

- כעת בברור הוא ידחה את השערת האפס.
- כעת הוא דווקא יקבל את השערת האפס.
- כעת לא ניתן לדעת מה תהיה מסקנתו.

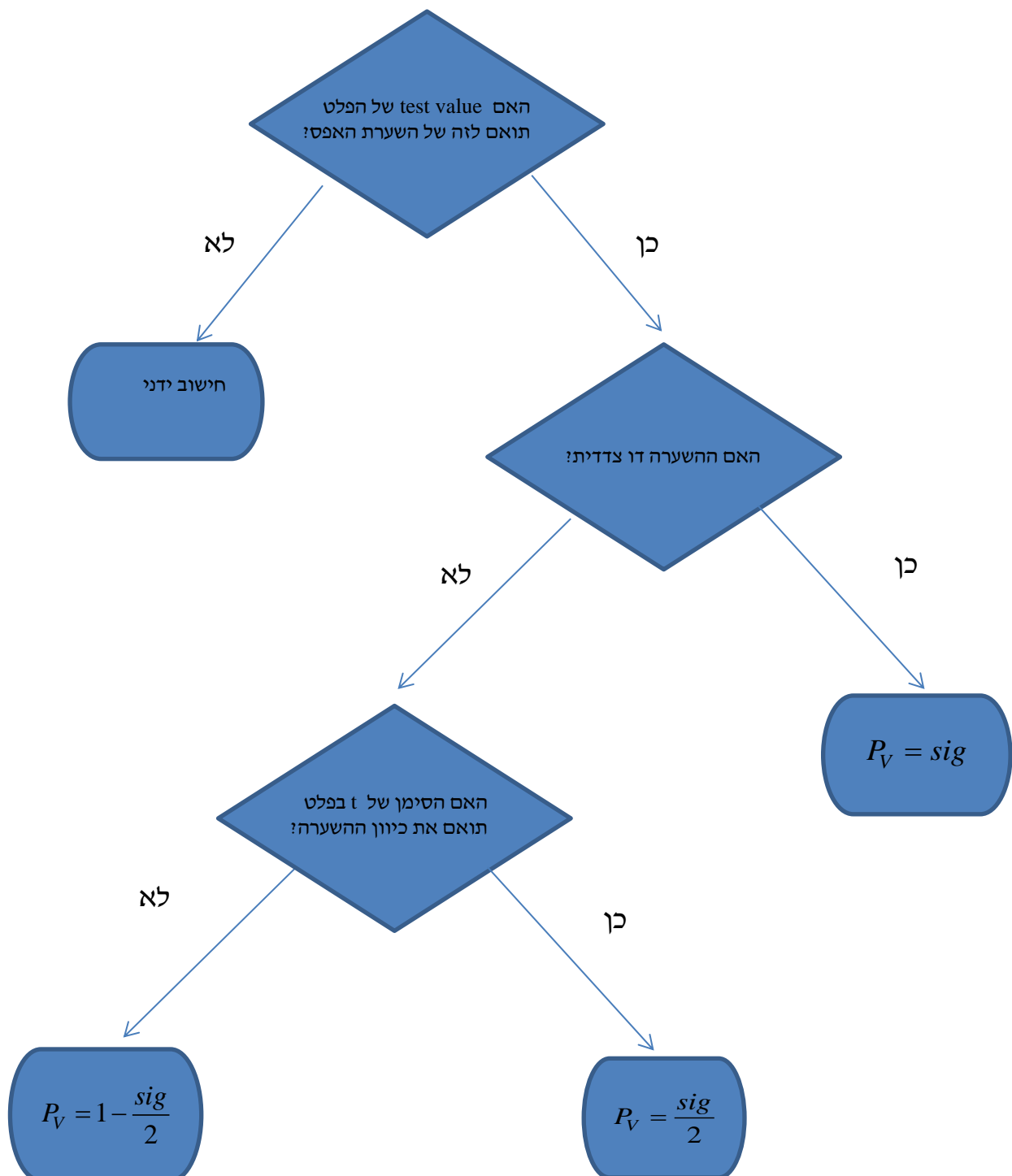
תשובות סופיות:

- (1) נדחה H_0 .
- (2) נדחה H_0 .
- (3) נקבל H_0 .
- (4) נקבל H_0 .
- (5) ב'.
- (6) ג'.

ניתוח פלטים:

רקע:

חישוב מובהקות התוצאה באמצעות פלט תוכנת SPSS:



דוגמה (פתרון בהקלטה):

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
X	25	87.6400	64.90434	12.98087

One-Sample Test

Test Value = 60						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
X	2.129	24	.044	27.64000	.8488	54.4312

ממוצע הציונים במבחן המיצב בחשבון הוא 60. הוחלט לדגום כיתה אקראית של 25 תלמידים וללמד אותם בשיטת לימוד חדשה.

- א. מהו רווח הסמך לממוצע הציונים בחשבון אם יוחלט ליישם את שיטת הלימוד החדשה?
- ב. מהו P_V לבדיקת יעילותה של שיטת הלימוד החדשה?
- ג. מה יוכרע ברמת מובהקות של 5% לגבי יעילותה של שיטת הלימוד החדשה?

שאלות:

- 1) באוניברסיטה גדולה גיל הסטודנטים לתואר ראשון מתפלג נורמאלי. בעבר פורסם שהגיל הממוצע של הסטודנטים הינו 23. להלן פלט תוכנת SPSS על מדגם של 16 סטודנטים אקראיים מתואר ראשון:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
age	16	23.4375	2.50250	.62562

One-Sample Test

	Test Value = 23					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
age	.699	15	.495	.43750	-.8960	1.7710

- א. מהו המבחן הסטטיסטי שנעשה כאן?
 ב. מה ערכו של הפרמטר לפי השערת האפס?
 ג. רשום רווח סמך ברמת סמך של 95% לתוחלת גיל הסטודנטים באוניברסיטה לתואר ראשון.
 ד. בדוק ברמת מובהקות של 5% האם הגיל הממוצע כיום שונה מבעבר?
- 2) קבוצת ילדים בגיל 6 קיבלה משימה לביצוע. עבור כל ילד בדקו כמה זמן לקח לו לסיים את המשימה בדקות. להלן תוצאות הניתוח הסטטיסטי:

One-Sample Test

	Test Value = 4.5					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
time	-1.853	24	.076	-.09200	-.1944	.0104

- א. כמה ילדים השתתפו במחקר?
 ב. מצא רווח סמך ברמת סמך של 95% לתוחלת זמן ביצוע המשימה עבור ילדים בני 6.
 ג. מה יש להניח כדי שרווח הסמך מסעיף א' יהיה תקף?
 ד. בדוק ברמת מובהקות של 5% שזמן ביצוע המשימה הממוצע נמוך מ-4.5 דקות.

3) להלן פלט מחשב עבור ניתוח סטטיסטי שנעשה בתוכנת SPSS. הניתוח הוא עבור מדגם אקראי של קבוצת נבחנים בבגרות באנגלית.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
grade	???	???	19.62787	2.95901

One-Sample Test

	Test Value = 75					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
grade	???	43	.017	-7.34091	-13.3083	???

- א. השלימו את הגדלים החסרים המסומנים בסמני שאלה בפלט.
 ב. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים שונה מ-75?
 ג. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים קטנה מ-75?
 ד. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים גדולה מ-75?

4) יצרן סיגריות מפרסם כי תוחלת הניקוטין בסיגריות שהוא מיצר קטנה מ-27 מ"ג. בבדיקה מקרית של 5 סיגריות מתוצרתו נמצאו כמויות הניקוטין הבאות: 21, 21, 20, 24, 22 מ"ג. הניחו כי כמות הניקוטין בסיגריות מפולג נורמאלי.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
nicotine	5	21.6000	1.51658	.67823

One-Sample Test

	Test Value = 27					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
nicotine	-7.962	4	.001	-5.40000	-7.2831	-3.5169

- א. האם ברמת מובהקות של 5% ניתן להסיק שיש אמת בפרסום?
 ב. אם היינו מוסיפים עוד תצפית שערכה 20. כיצד הדבר היה משפיע על הערך Sig ועל המסקנה?
 ג. בדקו האם ניתן להגיד שתוחלת רמת הניקוטין שונה מ-26 ברמת מובהקות של 5%.

תשובות סופיות:

- (1) א. הסקה של תוחלת אחת. ב. 23. ג. (22.104, 24.771). ד. נקבל H_0 .
- (2) א. 25. ב. (4.3056, 4.5104). ג. המשתנה מתפלג נורמלית. ד. נדחה H_0 .
- (3) א. $n = 44$, $\bar{X} = 67.66$, $t = -2.48$, $upper = -1.3735$. ב. 0.017. ג. 0.0085. ד. 0.9915.
- (4) א. נכריע שיש אמת בפרסום. ב. המסקנה לא תשתנה. ג. נכריע שהתוחלת שונה מ-26.

הקשר בין רווח סמך לבדיקת השערות על תוחלת (ממוצע):

רקע:

ניתן לבצע בדיקת השערות דו צדדית ברמת מובהקות α על μ :

$$H_0 : \mu = \mu_0 , H_1 : \mu \neq \mu_0$$

על ידי בניית רווח סמך ברמת סמך של $1-\alpha$ ל- μ :

אם μ_0 נופל ברווח \leftarrow נקבל את H_0 .

אם μ_0 לא נופל ברווח \leftarrow נדחה את H_0 .

דוגמה:

חוקר ביצע בדיקת השערות לתוחלת. להלן השערותיו :

$$H_0 : \mu = 80 , H_1 : \mu \neq 80 , \alpha = 5\%$$

החוקר בנה רווח סמך ברמה של 90% וקיבל: $79 < \mu < 84$.

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

פתרון (פתרון מלא בהקלטה):

רווח הסמך ברמת סמך של 90% מכיל "80".

ברמת סמך של 95% רווח הסמך יגדל ויכיל "80".

לכן, ברמת מובהקות של 5% נקבל H_0 .

שאלות:

- (1) חוקר רצה לבדוק את ההשערות הבאות: $H_0: \mu = 90$, $H_1: \mu \neq 90$. החוקר בנה רווח סמך לתוחלת ברמת סמך של 95% וקיבל את רווח הסמך הבא: (87, 97). אם החוקר מעוניין לבצע בדיקת השערות ברמת מובהקות של 1% האם ניתן להגיע למסקנה ע"ס רווח הסמך? נמקו.
- (2) חוקר מעוניין לבדוק השפעת דיאטה חדשה על רמת הסוכר בדם. ידוע כי מספר מיליגרם הסוכר בסמ"ק דם הוא משתנה מקרי שמתפלג נורמלית עם סטיית תקן 10.4 מ"ג. נלקח מדגם של 60 נבדקים שניזונו מדיאטה זו. נמצא כי ממוצע מספר המיליגרם סוכר היה 115.5 מ"ג לסמ"ק.
- א. בנה רווח סמך ברמת סמך 95% לתוחלת רמת הסוכר בדם אצל הניזונים מדיאטה זו.
- ב. ידוע שתוחלת רמת הסוכר בדם באוכלוסיה היא 90 מ"ג לסמ"ק. האם לדעתך ניתן להסיק על סמך תוצאת סעיף א שהדיאטה משפיעה על רמת הסוכר בדם? הסבירו.
- (3) יצרן אנטיביוטיקה רושם על גבי התרופות שכמות הפנצילין היא 200 מ"ג לקפסולה. משרד הבריאות ביצע מדגם של 8 קפסולות אקראיות מקו הייצור ומצא שבממוצע יש 196 מ"ג פנצילין לקפסולה עם סטיית תקן מדגמית של 5 מ"ג. בהנחה וכמות הפנצילין בקפסולה מתפלגת נורמלית.
- א. בנו רווח סמך ברמת סמך של 95% לממוצע כמות הפנצילין לקפסולה המיוצרת על ידי יצרן האנטיביוטיקה.
- ב. בדקו ברמת מובהקות של 5% האם יש אמת באינפורמציה המסופקת על ידי היצרן.

תשובות סופיות:

- (1) נקבל השערת.
- (2) א. $112.87 \leq \mu \leq 118.13$.
- ב. נכריע שהדיאטה משפיעה על תוחלת רמת הסוכר בדם.
- (3) א. $191.8 \leq \mu \leq 200.2$. ב. נכריע שיש אמת בפרסום.